PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N

A2

(11) International Publication Number: **WO 99/64562**

(43) International Publication Date: 16 December 1999 (16.12.99)

(21) International Application Number: PCT/US99/13266

(22) International Filing Date: 11 June 1999 (11.06.99)

(30) Priority Data:

60/089,049	12 June 1998 (12.06.98)	US
60/089,050	12 June 1998 (12.06.98)	US

(71) Applicant: THE PENNSYLVANIA STATE UNIVERSITY [US/US]; 113 Technology Center, 200 Innovation Boulevard, University Park, PA 16802–7000 (US).

(72) Inventors: GUILTINAN, Mark, J.; 801 Crabapple Court, State College, PA 16803 (US). KIM, Kyung–Nam; 6023 Avila Street, El Cerrito, CA 94530 (US).

(74) Agent: MONAHAN, Thomas, J.; The Pennsylvania State University, 113 Technology Center, 200 Innovation Boulevard, University Park, PA 16802–7000 (US).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: EXPRESSION CONTROL ELEMENTS FROM GENES ENCODING STARCH BRANCHING ENZYMES

(57) Abstract

Nucleic acid molecules for directing and modulating gene expression in transformed cells are disclosed. The molecules comprise 5' and 3' regulatory sequences isolated from genomic clones having homology in their coding regions with maize genes encoding Type I and Type IIB starch branching enzymes. Chimeric genes comprising these expression controlling nucleic acid molecules, and cells transformed with such chimeric genes, are also disclosed.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

EXPRESSION CONTROL ELEMENTS FROM GENES ENCODING STARCH BRANCHING ENZYMES

Pursuant to 35 U.S.C. §202(c), it is acknowledged that the U.S. Government has certain rights in the invention described herein, which was made in part with funds from the Department of Energy, Grant No. DE-FG02-96ER20234.

This application claims priority to U.S. Provisional Application Serial Nos. 60/089,049 and 60/089,050, filed June 12, 1998, the entireties of which are incorporated by reference herein.

FIELD OF THE INVENTION

The present invention relates to the field of genetic manipulation in plants. In particular, the invention provides a novel transcription and translation control elements isolated from genes encoding starch branching enzymes.

BACKGROUND OF THE INVENTION

Various publications or patents are referenced in this application to describe the state of the art to which the invention pertains. Each of these publications or patents is incorporated by reference herein.

Starch provides carbon and energy for vegetative and reproductive development of most higher plants. It is found as a water-insoluble granule which is mainly composed of two different polysaccharides, amylose and amylopectin. Amylose is considered to consist of linear α-1,4-linked glucose chains of about 1,000 residues long. However, amylopectin is a more highly branched macromolecule consisting of linear α-1,4-

glucose chains with α-1,6-glucosidic bonds at branch points.

Starch is synthesized in higher plants through the action of four classes of enzymes, ADP-glucose pyrophosphorylase (EC 2.7.7.23), starch synthase (EC 2.4.1.21), starch branching enzyme (EC 2.4.1.28) and starch debranching enzyme (EC 3.2.1.41). Starch branching enzymes (SBE) catalyze the formation of α-1,6 glucan linkages, thereby playing an important role in the synthesis of the amylopectin fraction of starch.

Multiple forms of starch branching enzymes (SBE) differing in enzymatic and biochemical properties have been identified and characterized in various plants. Three SBE isoforms, SBEI, IIa and IIb, have been isolated and characterized in maize. Isolation of the maize cDNAs encoding the SBE isoforms (*Sbe1*, *2a* and *2b*) enabled the investigation of the *Sbe* genes at molecular level (Fisher et al., Plant Physiol. 102: 1045-1046, 1993; Fisher et al., Plant Physiol. 108: 1313-1314, 1995; Gao et al., Plant Mol. Biol. 30: 1223-1232, 1996; Plant Physiol. 114: 69-78, 1997). Fisher et al. (Plant Physiol. 110: 611-619, 1996) determined that SBEIIa and IIb are the product of separate genes, and Gao et al. (1996, 1997, *supra*) demonstrated that the *Sbe* genes are differentially expressed during kernel development and in various tissues, suggesting that they play distinct roles in starch biosynthesis. For example, while *Sbe1* and *2a* are expressed in vegetative tissues, *Ae* is not. Moreover, unlike *Sbe1* and *2b*, *Sbe2a* is more highly expressed in embryos than in endosperm. However, all three genes are highly expressed in developing maize seed from mid to late development in both embryo and endosperm tissue.

Developmental and tissue specificity of gene

expression is governed by the expression regulatory regions of a gene. Typically, these include promoters, enhancers, 5' untranslated leaders, translation terminators and polyadenylation signals, among others. These *cis*-acting sequences control gene expression by regulating the timing, location and amount of transcription of the gene and/or stability or efficiency of translation of the encoded mRNA.

As maize seed is a major global food source, the ability to engineer protein production in seed will have many applications in agriculture and industry. High-level and tissue-specific expression regulatory sequences will be needed for the development of such genetically engineered plants. The expression regulatory sequences of the above-described *Sbe* and *Ae* genes from maize would be well suited for such applications, but heretofore have been unavailable.

SUMMARY OF THE INVENTION

According to one aspect of the present invention, an isolated nucleic acid molecule for controlling expression of genes in transformed plant cells is provided, which comprises a segment of an *Sbe1* gene from *Zea mays* or related species. In a preferred embodiment, the nucleic acid molecule is isolated from a gene having a coding sequence at least 60% homologous with the coding sequence defined by the exons of SEQ ID NO:1. In one embodiment, the segment begins at a location about 3,000 bases upstream from the transcription initiation site of the gene, and ends at a location about 250 bases downstream from the transcription initiation site. The molecule contains one or more specific regions for effecting high level

expression or sugar regulation. In another embodiment, the segment comprises a 3' untranslated region commencing at a stop codon for the gene's coding sequence, and ending at a location about 5900 bases downstream from the gene's transcription initiation site.

According to another aspect of the invention, a DNA segment for effecting expression of coding sequences operably linked to the segment is provided, which is isolated from a gene whose coding region hybridizes under stringent conditions with a coding region defined by exons of SEQ ID NO:1. Preferably, the gene is a maize *Sbe1* gene. In one embodiment, the segment comprises a promoter and a transcription initiation site, and may further comprise (1) an element included in the first exon of SEQ ID NO:1, which is capable of increasing promoter activity of homologous or heterologous promoters operably linked thereto, or (2) an element that confers sugar-regulatability on expression of the coding sequence. In another embodiment, the segment comprises a 3' untranslated region of the gene.

According to another aspect of the invention, another isolated nucleic acid molecule for controlling expression of genes in transformed plant cells is provided, which comprises a segment of a plant *Ae* gene. In a preferred embodiment, the nucleic acid molecule is isolated from a gene having a coding sequence at least 60% homologous with the coding sequence defined by the exons of SEQ ID NO:2, and most preferably it is a maize gene. In one embodiment, the segment begins at a location about 3,000 bases upstream from a transcription initiation site of the gene, and ends at a location about 100 bases downstream from the transcription initiation site. In another embodiment, the segment comprises a 3'

untranslated region commencing at a stop codon for gene's coding sequence, and ending at a location about 20,500 bases downstream from the gene's transcription initiation site.

In another aspect of the invention, a DNA segment for effecting expression of coding sequences operably linked to the segment is provided, which is isolated from a gene whose coding region hybridizes under stringent conditions with a coding region defined by exons of SEQ ID NO:2. Preferably, the gene is a maize *Ae* gene. In one embodiment, the segment comprises a promoter and a transcription initiation site. In another embodiment, the segment comprises a 3' untranslated region.

According to another aspect of the invention, a chimeric gene comprising a coding sequence operably linked to one or more of the aforementioned expression-regulatory sequences is provided. The chimeric gene preferably is inserted into a vector for transforming cells. Cells transformed with the vector are provided. In a preferred embodiment, they are plant cells, and are regenerated into fertile transgenic plants.

These and other features and advantages of the present invention will be described in greater detail in the description and examples set forth below.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1. Structure of the lambda clone 5-1-1 containing the *Sbe1* gene. **Fig. 1A.** Restriction map of the 5-1-1 clone. pBI5-1 and pBI5-2 indicate subclones in plasmid pBluescript SK$^-$. **Fig. 1B.** Genomic structure of the *Sbe1* gene. The thin black lines indicate the 5'- or 3'-flanking sequences of the *Sbe1* gene. The solid black

boxes indicate exons and the open boxes denote introns. The numbers represent positions relative to the transcription initiation site (+1).

Figure 2. Nucleotide sequence of the *Sbe1* gene and 5'- and 3'-flanking regions. Flanking regions and introns are shown in lowercase letters, while exons are presented in uppercase letters. The deduced amino acid sequences are shown below the string of exon sequences. Numbers indicate the distance relative to the transciption start site (+1) which is indicated by the arrowhead. The consensus TATA and G-box sequences as well as putative polyadenylation signal are underlined. The regions containing at least 82% sequence homology with the rice *Sbe1* 5' flanking region is also underlined. The asterisk and dot indicate the stop codon and putative polyadeylation site, respectively.

Figure 3. Genomic structure of the *Ae* gene. The complete structure of the *Ae* gene was constructed using two overlapping genomic clones: λ 3-2-1 and 7-2-1. The thick black lines indicate the 5'- or 3'-flanking sequences of the *Sbe1* gene. The solid —black boxes indicate exons and the open boxes denote introns. The position of a putative TATA-box relative to the transcription initiation site (+1) is indicated.

Figure 4. Nucleotide sequence of the *Ae* gene and 5'-and 3'-flanking regions. Numbers indicate the distance relative to the transciption start site (+1) which is indicated by the arrow. Intron sequences are omitted. Putative *cis*-elements are indicated below the sequence and boxed. The end points of the 5' deletion constructs were indicated by arrowheads below the first nucleotides of the deletions. Direct repeat sequences are underlined. The translation initiation codon (ATG)

and the translation termination codon (TGA) are double underlined. The dot indicates polyadenylation site. The region exhibiting 80% sequence similarity with the PREM-2 internal and 3' LTR region is boxed. The wavy line represents polypurine tract.

Figure 5. Schematic diagram of the 5' deletion chimeric constructs. **Fig. 5A.** The thick black lines denote the *Ae* promoter sequences. The numbers at left indicate deletion-end points relative to the transcription initiation site (+1) of the *Ae* gene. The open and stripped boxes indicate luciferase gene and nopaline synthase 3' end sequences, respectively. **Fig. 5B.** The junction sequences between the *Ae* gene and LUC. The BamHI site used to join the two genes is underlined. The translation start site of *LUC* is indicated by boldface letters. **Fig. 5C.** Effect of 5' deletions on *Ae* promoter activity. The relative activity values of the constructs are percentages of pKL201 level. Each value represents the average of three independent shootings. Error bars indicate standard errors of the means.

Figure 6. Effect of *Sbe1* gene exon/introns and 3' end on the level of LUC expression driven by the *Sbe1* promoter. **Fig. 6A.** Schematic diagram of chimeric *Sbe1* promoter-luciferase constructs. Numbers indicate distance relative to the *Sbe1* transcription start site. Translation initiation starts at a position, + 28. The light grey (stippled) boxes indicate the *Sbe1* promoter region. Angled lines indicate exons and introns in the *Sbe1* gene. Open and solid black boxes indicate luciferase (LUC) reporter gene and nopaline synthase 3' end sequences, respectively. The striped box indicates the *Sbe1* 3' flanking sequence. **Fig. 6B.** The junction sequences between the *Sbe1* gene and LUC. The BamHI sites

used to join the two genes are underlined. The translation start site of *LUC* is indicated by boldface letters. **Fig. 6C.** Expression levels of the construct shown in (A). LUC/GUS ratios were calculated as described in Methods. Each value represents the average of four independent shootings. Error bars indicate standard errors of the mean.

Figure 7. Effect of 5' deletions on *Sbe1* promoter activity. **Fig. 7A.** Schematic diagram of the 5' deletion chimeric constructs. The thick black lines denote the *Sbe1* promoter sequences. The numbers at left indicate deletion-end points relative to the transcription initiation site (+1) of the *Sbe1* gene. The light grey (stippled) boxes and the thin black angled line represent the first exon and intron in the *Sbe1* gene, respectively. The open boxes indicate the luciferase gene. The solid black boxes denote nopaline synthase 3' end sequences. **Fig. 7B.** The relative activity levels of the constructs shown in (A). The relative activity values are percentages of pKLN101 level. Each value represents the average of six to eight independent experiments. Error bars indicate standard errors of the means.

Figure 8. Linker-scan analyses of the 60-bp region in the *Sbe1* promoter. **Fig. 8A.** Schematic diagram of the linker-scan constructs. DNA sequence of the 60-bp region in the *Sbe1* promoter is shown to the right of the wild-type construct pKLN105. The mutated bases in the linker-scan constructs are shown in lowercase letters. Dashes represent the unaltered nucleotides. For an explanation of the other symbols, refer to the legend to Figure 13. **Fig. 8B.** Relative LUC activity levels of the constructs shown in (A). The relative activity values

are percentages of construct -314 level. Each value represents the average of four independent experiments. Error bars indicate standard errors of the means.

Figure 9. Sucrose responsiveness of the *Sbe1* promoter. Each construct was bombarded onto the maize endosperm suspension cell cultures supplemented with 0% (-) sucrose or 9% (+) sucrose and incubated for 48 hr at 25°C in the dark. The relative activity values are percentages of pKLN101 level in 0% sucrose. Each value represents the average of three independent experiments. Error bars indicate standard errors of the means.

Figure 10. Effect of mEmBP-1 overexpression on *Sbe1* promoter activity. 4 μg each of reporter plasmid (*Sbe1* promoter-LUC; pKLN101, or Ubiquitin-LUC; pACH18) and reference plasmid (CaMV 35S-GUS; pBI221) were co-precipitated onto gold particles with or without 4 μg of an effector plasmid (CaMV 35S-mEmBP-1). Maize endosperm suspension cells were bombarded wiht the gold particles and incubated at 25°C for 24 hr in the dark. The relative acvtivity values are percentages of the pKLN101 or pACH18 levels without mEmBP-1 overexpression. Each value represents the average of two independent shootings. Error bars indicate standard errors of the means.

DETAILED DESCRIPTION OF THE INVENTION

I. Definitions

Various terms used throughout the specification and claims to describe the invention. Unless otherwise specified, these terms are defined as set forth below.

With reference to nucleic acid molecules, the term "isolated nucleic acid" is sometimes used. This term, when applied to DNA, refers to a DNA molecule that

is separated from sequences with which it is immediately contiguous (in the 5' and 3' directions) in the naturally occurring genome of the organism from which it was derived. For example, the "isolated nucleic acid" may comprise a DNA molecule inserted into a vector, such as a plasmid or virus vector, or integrated into the genomic DNA of a procaryote or eucaryote. An "isolated nucleic acid molecule" may also comprise a cDNA molecule.

With respect to RNA molecules, the term "isolated nucleic acid" primarily refers to an RNA molecule encoded by an isolated DNA molecule as defined above. Alternatively, the term may refer to an RNA molecule that has been sufficiently separated from RNA molecules with which it would be associated in its natural state (i.e., in cells or tissues), such that it exists in a "substantially pure" form (the term "substantially pure" is defined below).

The term "substantially pure" refers to a preparation comprising at least 50-60% by weight the compound of interest (e.g., nucleic acid, oligonucleotide, protein, etc.). More preferably, the preparation comprises at least 75% by weight, and most preferably 90-99% by weight, the compound of interest. Purity is measured by methods appropriate for the compound of interest (e.g. chromatographic methods, agarose or polyacrylamide gel electrophoresis, HPLC analysis, and the like).

Nucleic acid sequences and amino acid sequences can be compared using computer programs that align the similar sequences of the nucleic or amino acids thus define the differences. For purposes of this invention, the GCG Wisconsin Package version 9.1, available from the Genetics Computer Group in Madison, Wisconsin, and the

default parameters used (gap creation penalty=12, gap extension penalty=4) by that program are the parameters intended to be used herein to compare sequence identity and similarity.

The term "substantially the same" refers to nucleic acid or amino acid sequences having sequence variation that do not materially affect the functionality of *cis* acting regulatory sequences (e.g, promoters, transcriptional response elements, etc.) or the nature of the encoded gene product (i.e. the structure, stability characteristics, substrate specificity and/or biological activity of the protein). With particular reference to nucleic acid sequences, the term "substantially the same" is intended to refer to conserved sequences governing expression and to the coding region (referring primarily to degenerate codons encoding the same amino acid, or alternate codons encoding conservative substitute amino acids in the encoded polypeptide).

The terms "percent identical" and "percent similar" are also used herein in comparisons among nucleic acid sequences. When referring to nucleic acid molecules, "percent identical" refers to the percent of the nucleotides of the subject nucleic acid sequence that have been matched to identical nucleotides by a sequence analysis program. When referring to amino acid sequences, "percent identical" refers to the percent of the amino acids of the subject amino acid sequence that have been matched to identical amino acids in the compared amino acid sequence by a sequence analysis program. "Percent similar" refers to the percent of the amino acids of the subject amino acid sequence that have been matched to identical or conserved amino acids. Conserved amino acids are those which differ in structure but are similar

in physical properties such that the exchange of one for another would not appreciably change the tertiary structure of the resulting protein. Conservative substitutions are defined in Taylor (1986, J. Theor. Biol. 119:205).

With respect to oligonucleotides or other single-stranded nucleic acid molecules, the term "specifically hybridizing" refers to the association between two single-stranded nucleic acid molecules of sufficiently complementary sequence to permit such hybridization under pre-determined conditions generally used in the art (sometimes termed "substantially complementary"). In particular, the term refers to hybridization of an oligonucleotide with a substantially complementary sequence contained within a single-stranded DNA or RNA molecule, to the substantial exclusion of hybridization of the oligonucleotide with single-stranded nucleic acids of non-complementary sequence.

A "coding sequence" or "coding region" refers to a nucleic acid molecule having sequence information necessary to produce a gene product, when the sequence is expressed.

The term "operably linked" or "operably inserted" means that the regulatory sequences necessary for expression of the coding sequence are placed in a nucleic acid molecule in the appropriate positions relative to the coding sequence so as to enable expression of the coding sequence. This same definition is sometimes applied to the arrangement other transcription control elements (e.g. enhancers) in an expression vector.

When describing the organization of a nucleic acid molecule, the term "upstream" refers to the 5' direction and the term "downstream" refers to the 3'

direction.

Transcriptional and translational control sequences, sometimes referred to herein as "expression control" sequences or elements, or "expression regulating" sequences or elements, are DNA regulatory elements such as promoters, enhancers, ribosome binding sites, polyadenylation signals, terminators, and the like, that provide for the expression of a coding sequence in a host cell. The term "expression" is intended to include transcription of DNA and translation of the mRNA transcript.

The terms "promoter", "promoter region" or "promoter sequence" refer generally to transcriptional regulatory regions of a gene, which may be found at the 5' or 3' side of the coding region, or within the coding region, or within introns. Typically, a promoter is a DNA regulatory region capable of binding RNA polymerase in a cell and initiating transcription of a downstream (3' direction) coding sequence. The typical 5' promoter sequence is bounded at its 3' terminus by the transcription initiation site and extends upstream (5' direction) to include the minimum number of bases or elements necessary to initiate transcription at levels detectable above background. Within the promoter sequence is a transcription initiation site (conveniently defined by mapping with nuclease S1), as well as protein binding domains (consensus sequences) responsible for the binding of RNA polymerase.

A "vector" is a replicon, such as plasmid, phage, cosmid, or virus to which another nucleic acid segment may be operably inserted so as to bring about the replication or expression of the segment.

The term "nucleic acid construct" or "DNA

construct" is sometimes used to refer to a coding sequence or sequences operably linked to appropriate regulatory sequences and inserted into a vector for transforming a cell. This term may be used interchangeably with the term "transforming DNA". Such a nucleic acid construct may contain a coding sequence for a gene product of interest, along with a selectable marker gene and/or a reporter gene.

A "heterologous" region of a nucleic acid construct is an identifiable segment (or segments) of the nucleic acid molecule within a larger molecule that is not found in association with the larger molecule in nature. Thus, when the heterologous region encodes a mammalian gene, the gene will usually be flanked by DNA that does not flank the mammalian genomic DNA in the genome of the source organism. In another example, coding sequence is a construct where the coding sequence itself is not found in nature (e.g., a cDNA where the genomic coding sequence contains introns, or synthetic sequences having codons different than the native gene). Allelic variations or naturally-occurring mutational events do not give rise to a heterologous region of DNA as defined herein.

A cell has been "transformed" or "transfected" by exogenous or heterologous DNA when such DNA has been introduced inside the cell. The transforming DNA may or may not be integrated (covalently linked) into the genome of the cell. For example, the transforming DNA may be maintained on an episomal element such as a plasmid. With respect to eukaryotic cells, a stably transformed cell is one in which the transforming DNA has become integrated into a chromosome so that it is inherited by daughter cells through chromosome replication. This stability is demonstrated by the ability of the

eukaryotic cell to establish cell lines or clones comprised of a population of daughter cells containing the transforming DNA. A "clone" is a population of cells derived from a single cell or common ancestor by mitosis. A "cell line" is a clone of a primary cell that is capable of stable growth *in vitro* for many generations.

II. Description

The genes encoding starch branching enzymes (SBEI, IIa and IIb) in maize are differentially regulated in tissue specificity and during kernel development. The expression-controlling sequences governing the differential regulation of genes encoding SBEI and SBEIIb are the subject matter of the present invention. These sequences are of great utility and value in genetic modification of plants to produce other gene products in the same tissue- and developmentally specific manner as are the two aforementioned starch branching enzymes. This is accomplished by placing the coding sequence of a gene of interest under the control of one or more of these expression-controlling sequences.

cDNAs for the starch branching enzymes have been disclosed, but such sequences, for the most part, do not contain the regulatory elements that control expression of a gene. To obtain these expression-controlling sequences, the actual genes must be isolated. In accordance with the present invention, the isolation and charcterization of two starch branching genes from maize, *Sbe1* and *Ae*, has enabled the identification of elements that regulate the expression of these genes.

The inventors have isolated and sequenced a maize genomic DNA (-2190 to +5929) which contains the entire coding region of SBEI (*Sbe1*) as well as 5'-and 3'-

flanking sequences. Sequence analysis of this gene (SEQ ID NO:1) is shown in Figure 2. Using this clone, the complete genomic organization of the maize *Sbe1* gene was established. The transcribed region consists of 14 exons and 13 introns, distributed over 5.7 kb. A consensus TATA-box and a G-box containing a perfect palindromic sequence, CCACGTGG, were found in the 5'-flanking region. Genomic Southern blot analysis indicated that two *Sbe1* genes with divergent 5'-flanking sequences exist in the maize genome, indicating that they may be differentially regulated. A chimeric construct containing the 5'-flanking region of *Sbe1* (-2190 to +27) fused to the β-glucuronidase gene (pKG101) showed promoter activity after it was introduced into maize endosperm suspension cells by particle bombardment. Although the 2.2-kb 5'-flanking sequence between -2190 and +27 relative to the transcription initiation site was sufficient to promote transcription, addition of the transcribed region between +28 and +228 containing the first exon and intron resulted in high-level expression in maize endosperm suspension cells. A series of 5' deletion and linker-substitution mutants identified two critical positive cis-elements, -314 to -295 and -284 to -255. Electrophoretic mobility shift assay showed that nuclear proteins prepared from maize kernels interact with the 60-bp fragment containing these two elements. Expression of the *Sbe1* gene is regulated by sugar concentration in cultured maize endosperm suspension cells, and the region -314 to -145 is essential for this effect. Expression of mEmBP-1, a bZIP transcription activator, in maize endosperm suspension cells resulted in a 5-fold decrease in the *Sbe1* promoter activity, suggesting a possible regulatory role of the G-box present in the *Sbe1* promoter

from -227 to -220.

A genomic clone of a *Sbe1* gene from rice has been described (Kawasaki et al., 1993, *supra*), as has an *Sbe1* promoter from wheat (U.S. Patent No. 5,866,793 to Baga, et al.) However, the presently described expression regulatory elements have several distinguishing features, described in greater detail below, which have not been reported for the rice or wheat promoter elements.

The *amylose-extender* (*Ae*) gene encoding SBEIIb in maize is predominantly expressed in endosperm and embryos during kernel development. To obtain the expression-controlling sequences of this gene, a maize genomic DNA fragment (-2,964 to +20,485) containing the *Ae* gene was isolated and sequenced. Sequence analysis of this gene (SEQ ID NO: 2) is shown in Figure 9. The maize *Ae* mRNA is derived from 22 exons distributed over 16,914 bp. Twenty one introns, differing in length from 76 bp to 4,020 bp, all have conserved junction sequences (GT . . AG). Sequence analysis of the 5'- and 3'-flanking regions revealed a consensus TATA-box sequence located 28 bp upstream of the transcription initiation site as determined by primer extension analysis, and a putative polyadenylation signal observed 29 bp upstream of the polyadenylation site based on cDNA sequence. Genomic Southern blot analysis revealed that a single *Ae* gene is present in the maize genome. Promoter activity was confirmed by testing a transcriptional fusion of the *Ae* 5'-flanking region between -2,964 and +100 to a luciferase reporter gene in a transient expression assay using maize endosperm suspension cultured cells. 5' deletion analysis revealed that the 111-bp region from -160 to -50 is essential for high-level promoter activity.

The present invention is drawn to all expression controlling elements of genes encoding SBEI and SBEIIB, particularly as exemplified by the maize *Sbe1* and *Ae* genes described in detail herein. With respect to the the maize *Sbe1* gene, the following sequences are considered particularly preferred for use in the present invention:

1. The 5' untranslated region, nucleotides -2190 to +27 relative to the transcription initiation site, which contains sequences sufficient to promote transcription;

2. The region containing the first exon and intron, nucleotides +28 to +228 relative to the transcription initiation site, which, when added to the promoter region, results in high level gene expression in the endosperm;

3. The region between -314 and -145 relative to the transcription initiation site, which is needed for sugar regulation of gene expression; and within this region

4. two critical positive cis-elements, -314 to -295 and -284 to -255; nuclear proteins prepared from maize kernels interact with the 60-bp fragment containing these two elements;

5. The 3' untranslated region, nucleotides +5390 to +5910, containing the polyadenylation signal and other sequences that may contribute to the stability or translation efficiency of the encoded mRNA.

With respect to the the maize *Ae* gene, the following sequences are considered particularly preferred for use in the present invention:

1. the 5' untranslated region, nucleotides -2,964 to +100 relative to the transcription initiation

site, which contains sequences sufficient to promote transcription, and within this region;

2. the 111-bp region from -160 to -50, which is essential for high-level promoter activity; and

3. The 3' untranslated region, nucleotides +16,695 to +20,485, containing the polyadenylation signal and other sequences that may contribute to the stability or translation efficiency of the encoded mRNA.

Although the genomic clones of the maize *Sbe1* and *Ae* are exemplified herein, this invention is intended to encompass nucleic acid sequences from other organisms, particularly higher plants, that are sufficiently similar to be used instead of the maize *Sbe1* and *Ae* nucleic acid moleucles for the purposes described below. These include, but are not limited to, allelic variants and natural mutants of SEQ ID NOS: 1 and 2 and the particular portions thereof listed above, which are likely to be found in different maize cultivars, as well as in different plant species. Of particular relevance to this invention are monocotyledenous plant species, and most especially those of agronomic importance, such as wheat, rice, barley and sorghum, for example. Also of particular relevance to this invention are taxonomic relatives of *Zea mays* (which includes cultivated maize and teosinte), including other members of the genus *Zea*, such as *Zea diploperennis* (diploperennial teosinte), *Zea luxurians* and *Zea perennis* (perennial teosinte), as well as *Tripsacum spp.* and *Sorghum* spp.

Variants from other cultivars and species are expected to possess certain differences in nucleotide sequence. Moreover, it is known that 5' and 3' regulatory regions of genes, while often sharing overall functional similarites, do not share a high degree of

sequence homology. Accordingly, this invention provides nucleic acid molecules comprising one or more the above-described expression regulatory sequences, which are isolated from genes with coding sequences (exons) having at least about 60% (preferably 70%, more preferably over 80% and most preferably over 90%) sequence homology with the respective coding sequences (exons)of the maize *Sbe1* or *Ae* genes, as exemplified by SEQ ID NO:1 or SEQ ID NO:2. The isolation of genes with exons having these levels of homology is accomplished by nucleic acid hybridization at a selected stringency, which is described in greater detail below.

It will also be understood by persons skilled in the art that the *Sbe1* and *Ae* expression-regulating elements of the invention include single- or double-stranded DNA or RNA.

The following description sets forth the general procedures involved in practicing the present invention. To the extent that specific materials are mentioned, it is merely for purposes of illustration and is not intended to limit the invention. Unless otherwise specified, general cloning procedures, such as those set forth in Sambrook et al., Molecular Cloning, Cold Spring Harbor Laboratory (1989) (hereinafter "Sambrook et al.") or Ausubel et al. (eds) Current Protocols in Molecular Biology, John Wiley & Sons (1999) (hereinafter "Ausubel et al.") are used.

Nucleic acid molecules comprising the promoters and/or other expression regulatory sequences of the invention may be prepared by two general methods: (1) they may be synthesized from appropriate nucleotide triphosphates, or (2) they may be isolated from biological sources. Both methods utilize protocols well

known in the art.

The availability of nucleotide sequence information, such as the aforementioned expression-controlling regions of SEQ ID NO:1 and SEQ ID NO:2, enables preparation of isolated nucleic acid molecules of the invention by oligonucleotide synthesis. Synthetic oligonucleotides may be prepared by the phosphoramadite method employed in the Applied Biosystems 38A DNA Synthesizer or similar devices. Variants of the aforementioned sequences also may be synthesized as described above.

Sbe1 and *Ae* genes and their expression-regulatory elements also may be isolated from appropriate biological sources using methods known in the art. In the exemplary embodiment of the invention, the genomic clones having SEQ ID NO:1 and SEQ ID NO:2 were isolated from a genomic library of maize inbred line B73. Genomic libraries of other maize cultivars or plant species, as discussed above, are also suitable sources for isolating these genes. A preferred means for isolating clones from a genomic library is PCR amplification using genomic templates and *Sbe1 or Ae*-specific primers derived from exons of SEQ ID NO:1 or SEQ ID NO:2.

In accordance with the present invention, genomic clones having the appropriate level sequence homology with the coding regions of SEQ ID NO:1 or 2 may be identified by using hybridization and washing conditions of appropriate stringency. For example, hybridizations may be performed, according to the method of Sambrook et al., using a hybridization solution comprising: 5X SSC, 5X Denhardt's reagent, 1.0% SDS, 100 μg/ml denatured, fragmented salmon sperm DNA, 0.05% sodium pyrophosphate and up to 50% formamide.

Hybridization is carried out at 37-42°C for at least six hours. Following hybridization, filters are washed as follows: (1) 5 minutes at room temperature in 2X SSC and 1% SDS; (2) 15 minutes at room temperature in 2X SSC and 0.1% SDS; (3) 30 minutes-1 hour at 37°C in 2X SSC and 0.1% SDS; (4) 2 hours at 45-55°in 2X SSC and 0.1% SDS, changing the solution every 30 minutes.

One common formula for calculating the stringency conditions required to achieve hybridization between nucleic acid molecules of a specified sequence homology (Sambrook et al., 1989):

$T_m = 81.5°C + 16.6\text{Log}[Na+] + 0.41(\%\ G+C) - 0.63\ (\%\ \text{formamide}) - 600/\#\text{bp in duplex}$

As an illustration of the above formula, using [N+] = [0.368] and 50% formamide, with GC content of 42% and an average probe size of 200 bases, the T_m is 57°C. The T_m of a DNA duplex decreases by 1 - 1.5°C with every 1% decrease in homology. Thus, targets with greater than about 75% sequence identity would be observed using a hybridization temperature of 42°C. Such a sequence would be considered substantially homologous to the sequences of the present invention. In a preferred embodiment, the hybridization is at 37°C and the final wash is at 42°C, in a more preferred embodiment the hybridization is at 42° and the final wash is at 50°, and in a most preferred embodiment the hybridization is at 42°C and final wash is at 65°C, with the above hybridization and wash solutions. Conditions of high stringency include hybridization at 42°C in the above hybridization solution and a final wash at 65°C in 0.1X SSC and 0.1% SDS for 10 minutes.

Nucleic acids of the present invention may be maintained as DNA in any convenient cloning vector. In a

preferred embodiment, clones are maintained in plasmid cloning/expression vector, such as pBluescript SK- (Stratagene, La Jolla CA) which is propagated in a suitable *E. coli* host cell.

The expression-controlling elements of the invention, as mentioned above, include transcriptional control elements such as promoters and enhancers that occur in flanking regions, in introns and sometimes in the coding region of the gene, as well as translational control elements (5' and 3') that affect the translation efficiency and/or stability of the mRNA. These elements are contemplated for use in a variety of applications in accordance with the present invention.

A preferred use of the expression-controlling elements is to drive transcription and translation of native and heterologous genes in transgenic plants, for the purpose of increasing the production of the endogenous gene products, or for producing new gene products in the tissue- and developmentally-specific manner governed by the expression-controlling sequences. Examples of useful gene products that can be expressed in seeds and other starch storage locations in plants include, but are not limited to: (1) gene products conferring herbicide tolerance; (2) gene products involved in starch synthesis, such as starch synthases or ADP-glucose pyrophosphorylase; (3) gene products involved in fatty acid biosynthesis, such as fatty acid desaturases; (4) gene products which are seed storage proteins, such as zeins; (5) gene products conferring resistance to insect pests, such as the crystal BT toxin; (6) gene products conferring resistance to microbial plant pathogens, including, e.g., viral coat proteins for virus resistance, or fungal cell wall lytic enzymes or

phytoalexins for fungal resistance; and (7) gene products that comprise or produce pharmaceutical or biological agents, such as antibiotics, secondary metabolites, antibodies, peptides or vaccines.

To use the expression-controlling sequences for introducing native or foreign genes into plants, the selected codings sequences and expression-controlling sequences are operably linked to one another, then inserted into vectors (or used directly, in some transformation protocols) for transforming plant cells. Methods for manipulating DNA sequences to accomplish this are well known in the art, as exemplified by Sambrook et al. and Ausubel et al. The design of such DNA constructs will depend on the expression functions desired for the gene of interest. As one example, all of the 5' and 3' expression-controlling sequences, as well as the +28 - +228 region, of the maize *Sbe1* gene may be operably linked to a coding sequence, such as a cDNA encoding a zein protein. A similar construct can be made using all of the 5' and 3' expression controlling elements of the *Ae* gene. In transgenic plants containing such chimeric genes, the zein protein would be produced in the same tissues under the same physiological conditions as the SBEI or SBEIIb proteins are produced. In another embodiment, if it is desired to produce a sugar-inducible gene, the segments of the 5' region of *Sbe1* controlling sugar regulation (-314 to -145) could be utilized independently.

In addition, certain of the expression regulatory elements (e.g., the -314 to -145 element or the +28 to +228 element of Sbe1) may be combined with constitutive or inducible promoters from other sources. Examples of constitutive promoters suitable for this

purpose are the rice actin promoter, the maize ubiquitin promoter and the Cauliflower Mosaic Virus (CaMV) 35S promoter. An example of a suitable inducible promoter is the tetracycline repressor/operator controlled promoter.

Other examples of ways to combine the various 5', 3' and internal expression-controlling elements of the *Sbe1* and *Ae* genes will be apparent to the person of skill in the art.

Transgenic plants expressing the native or heterologous chimeric genes described above can be generated using standard plant transformation methods known to those skilled in the art. These include, but are not limited to, *Agrobacterium* vectors, PEG treatment of protoplasts, biolistic DNA delivery, UV laser microbeam, gemini virus vectors, calcium phosphate treatment of protoplasts, electroporation of isolated protoplasts, agitation of cell suspensions with microbeads coated with the transforming DNA, direct DNA uptake, liposome-mediated DNA uptake, and the like. Such methods have been published in the art. See, e.g., Methods for Plant Molecular Biology (Weissbach & Weissbach, eds., 1988); Methods in Plant Molecular Biology (Schuler & Zielinski, eds., 1989); Plant Molecular Biology Manual (Gelvin, Schilperoort, Verma, eds., 1993); and Methods in Plant Molecular Biology - A Laboratory Manual (Maliga, Klessig, Cashmore, Gruissem & Varner, eds., 1994).

The method of transformation depends upon the plant to be transformed. The biolistic DNA delivery method is useful for nuclear transformation, especially of monocotyledenous plants. Methods for performing transformation using particle bombardment are described in detail in the Examples.

In another embodiment of the invention,

Agrobacterium vectors are used to advantage for efficient transformation of plant nuclei. In this embodiment, *Agrobacterium* "superbinary" vectors have been used successfully for the transformation of maize (Ishida et al., Nature Biotechnology 14: 745-750, 1996) and rice (Hiei et al., Plant Journal 6: 271-282, 1994).

Using an *Agrobacterium* superbinary vector system for transformation, the selected coding sequence under control of the expression-controlling elements of the invention, as described above, is linked to a nuclear drug resistance marker, such as kanamycin resistance or phosphothricin herbicide resistance. *Agrobacterium*-mediated transformation of plant nuclei is accomplished according to the following procedure:

(1) the gene is inserted into the selected *Agrobacterium* binary vector;

(2) transformation is accomplished by co-cultivation of plant tissue (e.g., leaf discs) with a suspension of recombinant *Agrobacterium*, followed by incubation (e.g., two days) on growth medium in the absence of the drug used as the selective medium;

(3) plant tissue is then transferred onto the selective medium to identify transformed tissue; and

(4) identified transformants are regenerated to intact plants.

Regardless of the transformation system used, it should be recognized that the amount of expression, as well as the tissue specificity of expression of the gene in transformed plants can vary depending on the position of their insertion into the nuclear genome. Such position effects are well known in the art. For this reason, several nuclear transformants should be regenerated and tested for expression of the transgene.

Another use of the expression-controlling elements is to inhibit or prevent transcription and translation of native *Sbe1* or *Ae* genes in plants. For this use, the elements are used as antisense molecules to block gene expression at critical points. Antisense molecules can be transiently expressed in plants by introduction as single-stranded molecules (DNA or RNA), according to known methods. Alternatively, vectors encoding the antisense molecules can be introduced into plants and the antisense molecules thereafter produced by transcription of the encoding sequences on the vectors, using the above-described methods. In another embodiment, overexpression of *Sbe1* or *Ae* is induced to generate a co-suppression effect. This excess expression serves to promote down-regulation of both endogenous and exogenous *Sbe1* or *Ae* genes.

The following specific examples are provided to illustrate embodiments of the invention. They are not intended to limit the scope of the invention in any way.

EXAMPLE 1
Genomic Organization and Promoter Activity of Maize Starch Branching Enzyme I Gene

This example describes the isolation of a full-length maize genomic DNA fragment containing the entire *Sbe1*gene. Structural and functional analysis of this gene revealed a complete genomic organization and demonstrated the transcriptional activity of its promoter region.

Materials and methods

PCR amplification. Maize (*Zea mays* L., inbred W64A) genomic DNA prepared from 22-DAP kernels using the method of Rogers and Bendich (1985) was amplified in a 50 μL reaction mixture containing 1 μg of target DNA, 1 μM of each primer, 200 μM of each dNTP, 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM $MgCl_2$, 0.01% gelatin (w/v), and 2.5 U of *Taq* DNA polymerase (Boehringer Mannheim, Indianapolis, IN). The mixture without dNTP was overlaid with 50 μL of mineral oil (Sigma, St.Louis, MO) and incubated for 5 min at 94°C. The dNTPs were added to the mixture at the end of the incubation and then the mixture was cycled 35 times in a thermal cycler (ERICOMP, San Diego, CA) as follows: 94°C for 30 s; 55°C for 1 min; and 72°C for 2 min; with a final 72°C extension of 7 min. The primers were designed according to the published sequence data of the maize partial *Sbe1* cDNA (Baba et al., Biochem. Biophys. Res. Commun. 181: 87-94, 1991). The 5' primer, 5'-GACTGAATTCCTGCGCAGGAGGCAGAGCTT-3' (SEQ ID NO:3), and the 3' primer, 5'-GATCGAATTCCATAGATACG TGGAGCAGCA-3' (SEQ ID NO:4), are homologous and complementary to DNA sequences of the maize *Sbe1* cDNA from 438 bp to 457 bp and 745 bp to 764 bp, respectively. Each primer contains an *Eco*R I restriction enzyme site and 4 extra nucleotides (underlined) at their 5' ends for convenience of subsequent cloning of the PCR product. After amplification, 15 μL of the reaction sample was run on an agarose gel (1.5 %,w/v) in 1X TAE buffer containing 0.04 M Tris-acetate and 1 mM EDTA. A single PCR product of about 0.5 kb was detected on an ethidium bromide-stained agarose gel, digested with *Eco*R I restriction enzyme and cloned into the corresponding site of pBluescript SK^- (Stratagene, La Jolla, CA) creating

plasmid pB1.

Maize genomic library screening and DNA sequencing. An EMBL-3 genomic library (Clonetech, Palo Alto, CA) prepared from maize seedlings (2-leaf stage, B73) was screened essentially according to Sambrook et al. (1989). Approximately 3 x 10^5 pfu were transferred onto nylon membranes (Hybond-N+, Amersham, UK) and hybridized with the ^{32}P-labeled *Sbe1* genomic PCR product excised from pB1. Hybridization was performed at 55°C for 20 h in 0.5 M Na_2HPO_4 (pH 7.2) and 7% SDS with gentle agitation at 40 cycles per minute on a rotary shaker. Following the hybridization, filters were washed twice in 5% SEN (5% SDS (w/v), 1 mM EDTA, 0.04 M Na_2HPO_4, pH 7.2) and once in 1% SEN (1% SDS, 1 mM EDTA, 0.04 M Na_2HPO_4, pH 7.2) for 15 min at 65°C. Plaques strongly hybridizing to the probe were selected and purified through three rounds of screening. Phage DNAs were isolated from the positive plaques according to Chisholm's method (1988) and were digested with *Pst* I to release the inserts from the EMBL-3 vector. The restriction fragments were separated on a 0.8% agarose gel and blotted onto nylon membranes. The blots were probed with ^{32}P-labeled full-length *Sbe1* cDNA (Fisher et al., 1995, *supra*) as described above and hybridizing DNA fragments were identified and subcloned into pBluescript SK-. DNA sequences were determined by the dideoxynucleotide chain termination method with Sequenase Version 2.0 (United States Biochemical Co., Cleveland, OH). Sequence analyses were performed using programs from DNASTAR Inc. (Madison, WI).

Primer extension analysis. To locate the transcription initiation site of the *Sbe1* gene, an oligonucleotide, 5'-GGCGACACGAGGCACAGCAT-3' (SEQ ID NO:5), which is complementary to the sense strand

sequence of the *Sbe1* cDNA from +1 to +20 relative to the translation start site (ATG) was radiolabeled at its 5' terminus with T4 polynucleotide kinase and γ^{32}P-ATP. Approximately 10^5 cpm of the labeled primer was hybridized at 35°C with 10 μg of total RNA, which was isolated from 30-DAP maize kernels (B73) according to the protocol of Vries et al. (Plant. Mol. Biol. Manual B6: 1-13, 1988). After hybridization for 8 h, complementary DNA was synthesized from the annealed primer by the addition of reverse transcriptase and dNTP. Following the addition of EDTA and RNAase A into the reaction, the nucleic acid was precipitated with ethanol. The reaction products were resuspended in sequencing gel loading buffer, denatured at 95°C, electrophoresed through a 5% polyacrylamide sequencing gel (w/v), and visualized by autoradiography. In order to provide size markers, part of the *Sbe1* gene was sequenced with the same primer used in the primer extension experiment.

3'-Rapid amplification of cDNA ends (3'-RACE). To isolate the 3' end of the *Sbe1* transcript, the 3'-RACE method was used (Frohman et al., 1988). First-strand cDNA synthesis reaction was performed as follows: 14 μL of a mixture containing 5 μg of total RNA from 30-DAP maize kernels (B73) and 50 pmol of a 39-bp oligonucleotide with 17 dT residues and an adaptor sequence, 5'-GGTCGACTCGAGTCGACATCGATTTTTTTTTTTTTTTT-3' (SEQ ID NO:6), was heated at 70°C for 10 min and quickly chilled on ice. To the chilled sample, 2 μL of 10X synthesis buffer (200 mM Tris-HCl, pH 8.4, 500 mM KCl, 25 mM $MgCl_2$, and 1 mg/mL BSA), 1 μL of 10 mM dNTP mix, 2 μL of 0.1 M DTT and 1 μL (200U) of SUPERSCRIPT reverse transcriptase (GIBCO BRL, Grand Island, NY) were added and incubated at room temperature for 10 min. The

reaction mixture was then placed in a 42°C water bath for 50 min and transferred to a 90°C water bath. After 5 min incubation, 1 μL of RNase H (2U/μL) was added and incubation at 37°C was continued for 20 min. Next, the first-strand cDNA obtained was amplified directly by the PCR method using a gene-specific primer (5'-GACTGAGCTCA TACCAAATGAAGCCAGGAG-3'(SEQ ID NO:7)), which is a homologous to sequence of the *Sbe1* gene from +5382 to +5401, and an adaptor primer (5'-GGTCGACTCGAGTCGACATCGA-3' (SEQ ID NO:8)). After amplification, a single DNA band detected on an agarose gel (1.5%, w/v) was isolated and digested with *Sac* I and *Xho* I (underlined within the primers). The resulting fragment, approximately 350 bp in length, was cloned into a pBlusScript SK⁻ (pB13R) and sequenced.

Genomic DNA blot analysis. Maize genomic DNA was prepared from 7-day-old etiolated seedlings (inbred B73) according to the method described by Junghans and Metzlaff (Junghans and Metzlaff, Biotechniques, 8: 176, 1990). 10 μg of genomic DNA was digested with restriction enzymes *Bam*H I, *Eco*R I, *Bgl* II, and *Hin*d III, separated on 0.8% agarose gels, and transferred onto nylon membranes (Hybond-N, Amersham, UK) in 20 X SSC containing 3 M NaCl and 0.3 M sodium citrate (pH 7.0) according to Sambrook et al. (1989). The DNA was crosslinked to the membrane by 3.5 min of UV irradiation on a transilluminator (312 nm). The genomic blots were prehybridized at 65°C for 1 h in 0.5 M Na_2HPO_4 (pH 7.2), 7% SDS, and 100 μg/mL denatured salmon sperm DNA. Using the random primed DNA labeling kit (Boehringer Mannheim, Indianapolis, IN), 25 ng of a full-length *Sbe1* cDNA (Fisher et al., 1995, *supra*) was labeled with [α-^{32}P]-dCTP. This labeled probe was added to the

prehybridization solution and incubated at 65°C for 18 h. Blots were washed twice in 5% SEN and once in 1% SEN for 15 min at 65°C and were exposed to Kodak X-AR film at -80°C for 48 h using two intensifying screens.

Construction of Sbe1 promoter-UidA expression plasmids. To make a transcriptional chimeric construct consisting of the *Sbe1* promoter (-2190 to +27) fused to a β-glucuronidase (GUS) reporter gene, *UidA*, a *Bam*H I restriction enzyme site was created just before the translation initiation site of the *Sbe1* gene as follows: the DNA sequence between -253 and +27 of the *Sbe1* gene was amplified via polymerase chain reaction using *Pfu* DNA polymerase (Stratagene, La Jolla, CA) to enhance the fidelity of PCR amplification. The 5' primer, 5'-CCAGCTCCACGGTTGTT CGTGT-3' (SEQ ID NO:9), is homologous to sequence of the *Sbe1* gene from -253 to -232. An *Apa* I restriction enzyme site (GGGCCC) is located immediately downstream of the 5'-primer-binding region of the *Sbe1* promoter, -203 to -198. The 3' primer, 5'-CGATGGATCC TGTGACGGCGTGTGAGT CCC-3' (SEQ ID NO:10), consists of a DNA sequence complementary to that of the *Sbe1* gene from +8 to +27 and a *Bam*H I restriction enzyme site (GGATCC) flanked with four random nucleotides (underlined). The PCR product was digested with *Apa* I and *Bam*H I, and the resulting 236-bp fragment was cloned into pBluescript SK$^-$ and sequenced in order to verify that no misincorporation had occurred in the DNA sequence during the PCR amplification. To incorporate this mutation into the context of a longer promoter fragment, the *Sbe1* genomic clone (λ 5-1-1) was digested with *Sal* I to isolate the approximately 3.0-kb fragment, which was then blunt-ended with Klenow and ligated with *Pst* I linkers (New England Biolabs). After complete digestion with *Pst* I and *Apa* I,

the resulting 2-kb *Sbe1* promoter fragment was ligated to the 236-bp *Apa* I-*Bam*H I fragment and cloned into plasmid pBI221 cut with *Pst* I and *Bam*H I, thereby creating plasmid pKG101.

Transient gene expression assay. Suspension culture cells of maize endosperm (inbred A636), kindly provided by J. L. Anthony (DEKALB Genetics Corporation, Mystic, CT) were grown in 250-mL large-mouth Erlenmeyer flasks containing 80 mL of Murashige and Skoog basal salt medium (Murashige and Skoog, 1962) supplemented with 0.4 mg/L thiamine, 2 g/L asparagine, and 30 g/L sucrose (Shannon and Liu, 1977). The culture was maintained in the dark at 29°C on a rotary shaker (120 rpm) and was subcultured every 7 days by transferring a portion of the cell suspension into fresh medium. For particle bombardment, the growing cells (3 days after subculture) were evenly distributed over the surface of three layers of filter paper (Whatmann #4, 55 mm in diameter) moistened with 3 mL of the liquid medium and positioned in the middle of a 10-cm petri dish. Three milligrams of 1.6-μm gold particles were coated with 10 μg of pKG101 according to the method described by Xu et al. (Xu et al., Plant Mol. Biol., 31: 1117-1127, 1996) and introduced into the cells using a Bio-Rad PDS-1000/He Biolistic Particle Delivery system. Bombardments were performed at 650 psi under a vacuum of 26 inches of Hg with a distance of 10 cm between the cells and the microprojectile launch site of the particle gun. Following the bombardments, the petri dishes were sealed with parafilm and then incubated in the dark at 25°C for 24 h.

Histochemical GUS staining. Histochemical assays of GUS activity were performed according to the

method described by Jefferson et al. (Jefferson et al., Plant Mol. Biol. Rep., 5: 387-405, 1987) with minor modifications. Briefly, after 24 h incubation, cells on the filter paper were transferred into petri dishes containing 2 mL of GUS staining solution (50 mM sodium phosphate buffer, pH 8.0, 0.05 mM potassium ferricyanide, 0.05 mM potassium ferrocyanide, 0.2% Triton X-100 (w/v), 20% methanol (v/v), and 1 mg/mL 5-bromo-4-chloro-3-indolyl-β-D-glucuronide). The petri dishes were then sealed and incubated at 37°C overnight. The cells were examined and photographed under a Nikon SMZ-U dissecting microscope.

Results

PCR amplification of maize genomic DNA. In order to obtain a DNA probe for genomic library screening, maize genomic DNA prepared from 22-DAP kernels (W64A) was amplified by polymerase chain reaction (PCR) using upper and lower primers designed to anneal the *Sbe1* cDNA (Baba et al., Biochem. Biophys. Res. Commun., 181: 87-94, 1991) from 438 to 457 and 745 to 764, respectively. A single amplified DNA band, approximately 450 bp in length, was observed on an ethidium bromide-stained agarose gel. The PCR product containing *Eco*R I sites at both 5'- and 3'-ends was digested with *Eco*R I and the resulting fragment was cloned and sequenced.

Alignment of sequences between the maize PCR genomic fragment and the published *Sbe1* cDNA (Baba et al., 1991, *supra*) shows that the PCR fragment contained the predicted *Sbe1* cDNA sequence with only 3 nucleotide differences and a 103-bp intron. The differences could be explained by cultivar polymorphisms or misincorporation of bases during PCR amplification by *Taq*

DNA polymerase which does not have proofreading activity. This PCR fragment was ^{32}P-labeled and used as a hybridization probe for screening a maize genomic library to isolate *Sbe1* genomic clones.

Isolation and analysis of a maize Sbe1 genomic clone. In a screen of approximately 3 x 10^5 plaque-forming units from a genomic library prepared from maize seedlings (inbred B73), 8 positive lambda clones were isolated which strongly hybridized to the probe. Restriction mapping and partial DNA sequencing of these clones indicated they all probably originated from the same genetic locus. A full-length clone (λ 5-1-1), containing the entire coding region of the *Sbe1* gene, as well as 5'- and 3'- flanking sequences, was selected for further analyses. Figure 1A shows a restriction map of the genomic clone. The 3.0-kb *Sal* I fragment and the 6.2-kb *Pst* I fragment from the clone were subcloned into pBluescript SK$^-$ producing plasmids pBI5-1 and pBI5-2, and their nucleotide sequences were completely determined.

A consensus TATA-box as well as a G-box containing a perfect palindromic sequence known as a G-box, CCACGTGG, were found in the 5' flanking region of the gene (Fig. 2). Primer extension analysis displayed a single extended product of 44 nucleotides, which co-migrates with an A residue in the sequencing ladder generated with the *Sbe1* 5'-flanking region. This indicates that consistant with many eukaryotic genes, the transcription initiates at a position which is located 25 bp downstream from the putative TATA box, suggesting that the TATA box may be a functional element of the *Sbe1* promoter.

To determine the polyadenylation site of the *Sbe1* gene, 3' RACE was conducted. It demonstrated that a

poly (A) tail occurs at an adenine nucleotide located 29 bp downstream from a putative polyadenlyation signal (AATAAA) in the *Sbe1* gene. Along with the primer extension result, this indicated that the transcribed region in the *Sbe1* gene is 5,690 bp in length.

Structure of the Sbe1 genomic clone. Alignment of the genomic sequence with the published maize *Sbe1* cDNA sequence (Fisher et al., 1995, *supra*) revealed that the gene is composed of 14 exons and 13 introns distributed over 5.7 Kb. Figure 1B summarizes the organization of the maize *Sbe1* gene. The cDNA sequence is identical to the corresponding genomic sequence except for a 1-bp mismatch in exon 14. Table 1 shows the sequences around the exon/intron junctions and a list of putative branch point consensus sequences, which were derived as described by Brown (Brown, J.W.S., Nucl. Acids Res., 14: 9549-9559, 1986).

Table 1. List of introns and sequences of exon/intron borders in the *Sbe1* gene

Intron number	Exon /	Putative intron branch point*	/Exon	Intron size(bp)	GC content (%)
1	TCGCG	GTAAG ···· CTGAT··17[Y]·CGCAG	GGTGG	82	46.3
2	GGAAG	GTAGA ···· CTGAA··24··GGAAG	GTCAA	377	36.9
3	TAAAG	GTTAG ···· ATCAT··30··TTCAG	GCTAT	120	34.2
4	GCGCA	GTAAG ···· TTGAC··25··TGTAG	GGAGG	319	34.2
5	GAAAG	GTCTC ···· ATGAC··33··TGCAG	GTACA	103	35.9
6	ATCAG	GTACC ···· ATGAC··43··TTCAG	TCTAT	262	32.1
7	AAAAG	GTTCC ···· CTCAA··33··TCCAG	ATGAT	97	39.2
8	ATGAG	GTGAA ···· TTGAT··17··TCTAG	TTTGG	488	35.2
9	ACAAG	GTTAT ···· CTAAC··37··AACAG	TACAT	565	37.3
10	AAAAG	GTAAG ···· GTCAG··25··TTCAG	GTTAT	160	39.4
11	GAGGG	GTAAG ···· CTTAC··31··TGCAG	CTACA	107	41.1
12	GAAGA	GTAAG ···· CTCAT··16··CGCAG	GTTGG	140	49.3
13	GTGTG	GTAAT ···· CTGAC··18··GCCAG	GCTTA	73	49.3

* Consensus sequences between introns are underlined.
[Y] Numbers indicate number of nucleotides between adjacent sequences.

The introns, relatively AT-rich (61%) compared to the exons (52%), vary in length from 73 bp to 565 bp, and all of which have the conserved sequences at their 5' and 3' ends, following the 'GT . . AG' rule of plant introns (Brown, 1986, *supra*). Exon 1, containing 27 bp of 5'-untranslated DNA sequence, and exon 2 occur in the transit peptide region which may be essential for transporting the gene product into the amyloplast. Exon 14 contains the translation stop codon (TGA) and 3'-untranslated region, as well as the putative polyadenylation signal (AATAAA). The exons vary in length from 63 to 907 bp.

Comparison of the maize and rice *Sbe1* genomic DNA sequences (Kawasaki et al., Mol. Gen. Genet., 237: 10-16, 1993b) revealed two large highly conserved regions in the 5'-flanking sequences (Fig. 2). One region, 161 bp in length, was present between -2190 and -1890 in the maize *Sbe1* 5'- flanking sequence, and has 82% similarity with the corresponding rice region. The other, 342 bp in length, is located from -1804 to -1611, and shows 85% similarity with the corresponding rice region. These sequences conserved between the species suggest that the regions may play an important role in gene expression. Interestingly, a part of the latter region from -1804 to -1611 (194 bp in length) shares 83% similarity with the 5'-untranslated sequence of a Ca^{2+}-dependent protein kinase gene in rice (Kawasaki et al., Gene., 129: 183-189, 1993a), suggesting that a maize version of protein kinase gene may be located immediately upstream of the *Sbe1* gene. In the rice genome, the two genes are separated by approximately 1.4 kb, and transcribed divergently from each other.

5'-flanking sequences downstream from the two

conserved regions did not show sequence similarity between the two genes (less than 25%) except for a G-box motif and several small segments adjacent to the G-box. This may indicate that the G-box plays a role in regulation of *Sbe1* gene expression. The G-box is found in other plant genes which respond to diverse environmental or physiological stimuli and are often associated with additional regions which possibly act as coupling elements determining signal response specificity.

Another notable feature derived from the sequence comparison is that the maize and rice *Sbe1* genomic structures are quite similar. Both genes consist of 14 exons and 13 introns, the positions of which are conserved between the two species. The sizes of exons (exon 3 to exon 13) constituting most of the mature proteins are identical except exon 5, in which the rice *Sbe1* gene has one more codon compared to the maize gene. They share more than 86% and almost 90% similarity in nucleotide and amino acid sequences, respectively. However, exons not encoding the mature proteins (exon 1, 2 and most of exon 14) do not display any significant sequence similarity and vary in size. The carboxyl-terminal 64 (67 in rice) amino acids encoded by exon 14 in the maize gene is the only region which is not conserved in the mature protein. Unlike the exons, homology was not found in any of the introns other than the splice junction sequences. The large intron (2212 bp) present in the rice *Sbe1* gene is not found in the maize gene.

Genomic Southern blot analysis. To determine the number of *Sbe1* genes in the maize genome, Southern blot analysis was performed. When blots were hybridized

with the full-length maize *Sbe1* cDNA (Fisher et al., 1995, *supra*) under high-stringency conditions, at least three bands were detected in each lane. Comparison of the hybridization patterns with the restriction map of the *Sbe1* genomic clone (Fig. 1A) revealed that not all the bands in the Southern blot corresponded to the genomic map, suggesting more than one *Sbe1* gene is present in the maize genome.

To confirm this, a 0.6-kb genomic DNA probe which did not have any restriction enzyme sites used in the genomic blot was prepared from the genomic clone λ 5-1-1 by *Bam*H I-*Hin*d III digestion. The genomic probe will produce only one hybridizing band in every lane if there is a single copy of the *Sbe1* gene in the maize genome. However, the genomic probe (probe 2) containing the central region of the *Sbe1* cDNA detected in each lane one or two additional bands apart from the bands predicted by the genomic map. This indicates that along with the isolated *Sbe1* gene, another *Sbe1* gene or a gene very closely related to *Sbe1* exists in the maize genome. These will be referred to as *Sbe1a* (identified) and *Sbe1b* (unidentified) to distinguish them when appropriate. Interestingly, when a 1.7-kb *Bam*H I-*Pst* I genomic fragment consisting of the *Sbe1a* promoter and transit peptide-coding region was used as a probe (probe 1), only the bands predicted from the identified *Sbe1a* genomic DNA sequence were detected. Taken together, these results suggest that although two *Sbe1* genes are present in the maize genome, their 5' flanking sequences and the 5'-end of the coding regions (at least in DNA level) are quite divergent from each other.

Genetic mapping of Sbe genes. In order to map *Sbe1* loci onto the existing framework map (Gardiner et

al., Genetics, 134: 917-930, 1993), the full length *Sbe1* cDNA was used to probe a population of 54 immortalized F_2 individuals from the cross Tx303 X CO159 as part of the mapping efforts of the University of Missouri-Columbia Maize RFLP Laboratory. Two loci were identified on chromosome 6, bin 6.01 and chromosome 10, 10.04. This supports the conclusion made above based on genomic Southern analysis that two *Sbe1* genes exist in the maize genome. However, at present the data do not allow us to distinguish which gene is located on which chromosome.

Promoter activity. A transient expression system was used to test whether the 5'-flanking sequence of the cloned *Sbe1* gene (λ 5-1-1) is sufficient to support transcriptional activity. A chimeric gene containing a 2.2-kb 5' *Sbe1* fragment (-2191 to +27) fused to *UidA* reporter gene in pUC19 was first constructed and designated pKG101. The chimeric plasmid was introduced into maize endosperm suspension cells via particle bombardment. Iodine staining and northern blot analysis showed that the maize endosperm suspension cells actually produce starch and the genes involved in starch biosynthesis are expressed. The bombarded cells were incubated for 24 h at 25°C in the dark and histochemical GUS assays were performed to visualize GUS expression. Blue spots were observed, indicating that the 5'-flanking sequence has the ability to drive gene expression in the maize endosperm cells. Control cells bombarded with gold particles coated with promoterless *UidA* construct did not show any blue spots.

EXAMPLE 2
Molecular Cloning and Characterization of the *Amylose-Extender* Gene Encoding Starch Branching Enzyme IIB in Maize

This example describes the complete genomic organization of the *Ae* gene and the promoter regions critical for its expression in maize endosperm cells.

Materials and Methods

Maize Genomic Library Screening and DNA Sequencing. Using ^{32}P-labeled full-length *Ae* cDNA (Fisher et al., 1993, *supra*) as probe, an EMBL-3 genomic library (Clonetech, Palo Alto, CA) prepared from maize seedlings (B73, 2 leaf stage) was screened essentially according to Sambrook et al. Approximately 3 x 10^5 plaque forming units were transferred onto nylon membranes (Hybond-N+, Amersham, UK). Hybridization was performed at 55°C for 20 h in a solution containing 0.5 M Na_2HPO_4, pH 7.2, and 7% SDS with gentle agitation at 40 cycles per minute on a rotary shaker. Following the hybridization, filters were washed twice in 5% SEN (5% SDS (w/v), 1 mM EDTA, 0.04 M Na_2HPO_4, pH 7.2) and once in 1% SEN (1% SDS, 1 mM EDTA, 0.04 M Na_2HPO_4, pH 7.2) for 15 min at 65°C. Plaques strongly hybridizing to the probe were selected and purified through three rounds of screening. Phage DNAs were isolated from the positive plaques according to Chisholm's method and SalI-digested inserts were subcloned into pBluescript SK$^-$. DNA sequences were determined by the dideoxynucleotide chain termination method with Sequenase Version 2.0 (United States Biochemical Co., Cleveland, OH). Sequence analyses were performed using programs from DNASTAR Inc. (Madison, WI).

Primer Extension Analysis. To locate the transcription initiation site of the *Ae* gene, an oligonucleotide, 5'-GATCGGATCGAACTGATCAG-3' (SEQ ID NO:11), which is complementary to the sense strand sequence of the *Ae* cDNA from -35 to -16 relative to the translation start site (ATG) was radiolabeled at its 5' terminus with T4 polynucleotide kinase and γ^{32}P-ATP. Approximately 10^5 cpm of the labeled primer was hybridized at 35°C with 10 μg of total RNA, which was isolated from 30 DAP maize kernels (B73) according to the protocol of Vries et al. (In: Gelvin, SB, Schilperoot, RA (eds) Plant Molecular Biology Manual B6: 1-13, Kluwer Academic Publishers, Dordrecht, Netherlands, 1988). After hybridization for 8 h, complementary DNA was synthesized from the annealed primer by the addition of reverse transcriptase and dNTP. Following the addition of EDTA and RNAase A into the reaction, the nucleic acid was precipitated with ethanol. The reaction products were resuspended in sequencing gel loading buffer, denatured at 95°C, electrophoresed through a 5% polyacrylamide sequencing gel (w/v), and visualized by autoradiography. In order to provide size markers, part of the *Ae* gene was sequenced with the same primer used in the primer extension experiment.

Genomic Southern Blot Analysis. Maize genomic DNA was prepared from 7-day-old etiolated seedlings (inbred B73) according to the method described by Junghans et al (1990, *supra*). 10 μg of genomic DNA was digested with restriction enzymes, separated on 0.8% agarose gels, and transferred onto nylon membranes (Hybond-N, Amersham, UK) in 20 X SSC containing 3 M NaCl and 0.3 M sodium citrate, pH 7.0 according to Sambrook et al. DNA was crosslinked to the membrane by 3.5 min of UV

irradiation on a transilluminator (312 nm). Genomic blots were prehybridized at 65°C for 1 h in 0.5 M Na_2HPO_4, pH 7.2, 7% SDS, and 100 μg/mL denatured salmon sperm DNA. 25 ng of a full-length *Ae* cDNA (Fisher et al., 1993, *supra*) was labeled with [α-^{32}P]-dCTP using a random primed DNA labeling kit (Boehringer Mannheim, Indianapolis, IN). Labeled probe was added to the prehybridization solution and incubated at 65°C for 18 h. Blots were washed twice in 5% SEN and once in 1% SEN for 15 min at 65°C and were exposed to Kodak X-AR film at -80°C for 48 h using two intensifying screens.

Construction of Plasmids. For a transcriptional fusion of the *Ae* promoter to a luciferase (LUC) reporter gene, a BamHI restriction enzyme site was created just before the translation initiation site (ATG) of the *Ae* gene as follows. The DNA sequence between -14 and +100 of the *Ae* gene was first amplified via polymerase chain reaction. The primer (PII-2), 5'-CCTAATTGTAGCCCTGCAGTCA-3' (SEQ ID NO:12), is homologous to sequence of the *Ae* gene from -10 to +12. A PstI site (CTGCAG) is located immediately downstream of the primer binding region of the *Ae* promoter, +4 to +9. The 3' primer (PII-3), 5'-GACTGGATCCTCGCCTTCGCAGCCGGATCG-3' (SEQ ID NO:13), consists of a DNA sequence complementary to that of the *Ae* gene from +80 to +100 and a BamHI restriction enzyme site (GGATCC) flanked with four random nucleotides (underlined). The PCR product was digested with PstI and BamHI, and the resulting 100-bp fragment was ligated to the 2,977-bp SalI-PstI *Ae* promoter fragment and cloned into plasmid pLN cut with SalI and BamHI (promoterless LUC-NOS gene in pUC119), thereby creating plasmid pKL201. This construct as well as all the following constructs

were verified by DNA sequencing.

To construct a translational fusion of the *Ae* promoter containing the first exon and intron to a LUC reporter plasmid, the *Ae* genomic clone, 3-2-1, was digested with XhoI and the resulting 866-bp fragment was gel purified. The fragment was blunt ended by Klenow fill-in DNA synthesis and ligated with BamHI linkers (CGGGATCCCG). After complete digestion with PstI and BamHI, the 325-bp DNA fragment was isolated and used to replace the 100-bp PstI-BamHI region in pKL201. This construct was designated pKLN201.

A series of 5' deletion mutants were derived from pKL201 using available restriction enzyme sites and PCR techniques. To create pKL202, pKL203, pKL205 and pKL206, pKL201 was first digested with AccI (-1719 to -1714), SpeI (-1128 to -1123), XhoI (-537 to -532) and ApaI (-348 to -343), respectively. Then, each linearlized plasmid was separately gel purified and blunt ended by Klenow fragment. After SalI linker ligation, the modified plasmids were digested with SalI, and the larger DNA fragments from each reaction were isolated and self-ligated to produce the relevant plasmids carrying different deletion end points.

For a construction of pKL204, the *Ae* promoter region between -755 and -500 was amplified with two primers. The 5' primer containing a SalI (GTCGAC) flanked with four extra nucleotides (underlined), 5'-GAAAGTCGACGAAGAGAGAATGAAAGCGAA-3'(SEQ ID NO:14), and the 3' primer, 5'-GCGCGGGTCCGTCGTGCCTTTT-3'(SEQ ID NO:15), were designed to anneal to DNA sequences of the *Ae* gene from -755 to -736 and from -521 to 500, respectively. The amplified 266-bp product was digested with SalI and XhoI which was located 10 bp upstream of the 3' primer

binding region, and the resulting 230-bp fragment was used to substitute the 2.4-kb SalI-XhoI fragment in pKL201.

To make pKL207, the *Ae* promoter region between -160 and +100 was first amplified with 5' primer, 5'-GATAGTCGACCGACGCGCAACGGCCTGCCT-3' (SEQ ID NO:16), and 3' primer, 5'-GACTGGATCCTCGCCTTCGCAGCCGGATCG-3' (SEQ ID NO:17), which contain SalI and BamHI sites, respectively, along with four arbitrary extra bases (underlined). Next, the PCR product was digested with SalI and BamHI and the resulting 267-bp DNA fragment was then used to replace the 3.1-kb SalI-BamHI fragment in pKL201. The same 3' primer and method were used for a construction of pKL208 except for the 5' primer, 5'-GATCGTCGACCGCTCGTCTCCGTCCTATAT-3' (SEQ ID NO:18), homologous to the DNA sequence of the *Ae* gene from -49 to -32.

Particle Bombardment. Suspension culture cells of maize (inbred A636) endosperm provided by J. L. Anthony (DEKALB Genetics Corporation, Mystic, CT) were grown in 250-ml large-mouth Erlenmeyer flasks containing 80 ml of Murashige and Skoog basal salt medium supplemented with 0.4 mg/l thiamine 2 g/l asparagine and 30 g/l sucrose. The culture was maintained in the dark at 29°C on a rotary shaker (120 rpm) and was subcultured every 7 days by transferring a spoonful of the cell suspension into 80 ml of fresh medium.

For particle bombardment, about 600 mg (fresh weight) of actively growing cells 3 days after subculture was evenly distributed over the surface of filter paper (Whatmann #4, 55 mm in diameter) by vacuum filtration of 8 ml of suspension culture. The filter paper bearing the cells was then placed over three layers of filter paper

(Whatmann #4, 70 mm in diameter) moistened with 5 ml of the liquid medium containing 12% sucrose and positioned in the middle of a 10 cm petri dish.

60 mg of gold microcarriers (1.6 μm particle size) were washed three times with 1 ml of 100% ethanol and twice with 1 ml of sterile deionized H_2O, resuspended in 1 ml of sterile deionized H_2O, and dispensed in 50-μl aliquots (3 mg/50 μl). The *Sbe1* Promoter-LUC constructs and a GUS reference plasmid (pBI 221; Clontech, Palo Alto, CA) were co-precipitated onto gold particles as follows: under continuous vortexing, the following were added in order to each 50-μl aliquot of gold particles: 5 μl of DNA (8 μg of LUC reporter plasmid, 4 μg of GUS reference plasmid), 50 μl of 2.5 M $CaCl_2$, and 20 μl of 0.1 M spermidine (free base, tissue culture grade). The gold particles coated with DNA were pelleted in an Eppendorf centrifuge at 10,000 rpm for 10 sec, rinsed with 250 μl of 100% ethanol, and resuspended in 60 μl of 100% ethanol. Immediately after sonication, 8 μl of the DNA-coated gold particles were pipetted onto the center of macrocarriers (Bio-Rad, Hercules, CA) and dried in a low humidity environment.

A Bio-Rad PDS-1000/He Biolistic Particle Delivery system was used for particle bombardment. Bombardment parameters which were optimized include He pressure, gap distance (distance from power source to macroprojectile), and target distance (distance from microprojectile launch site to sample target). After optimization, all bombardments were performed in a dim room at 650 psi under a vacuum of 26 inches of Hg with a distance of 10 cm between the cells and the barrel of the particle gun. Following the bombardments, the petri dishes were sealed with Parafilm and then incubated in

the dark at 25°C for 24 hr.

GUS and LUC Assays. The bombarded cells were harvested from the plates by vacuum filtration, frozen in liquid nitrogen, and ground with a pestle and mortal to a fine powder. The powder was then transferred into a microfuge tube and extracted with cell culture lysis buffer containing 300 mM Tris-phosphate, pH 7.8, 2 mM DTT, 2 mM 1,2-diaminocyclohexane-N, N, N', N'-tetraacetic acid, 10% glycerol and 1% Triton X-100 (0.3 ml/g of tissue). Cell debris was pelleted in an Eppendorf centrifuge at 14,000 rpm for 10 min at 4°C and the supernatant was split into two aliquots for assays of GUS and LUC activity.

For fluorometric GUS assays, 30 μl of the crude extract was incubated at 37°C with 2 mM 4-methyl umbelliferyl glucuronide in 0.3 ml of GUS assay buffer (50 mM $NaPO_4$, pH 7, 10 mM EDTA, 0.1% Triton X-100, 0.1% Sarkosyl, 10 mM β-mercaptoethanol, 20% methanol). After 0, 1, and 2 hr of incubation, 0.1 ml aliquots were removed and added to 0.9 ml of 0.2 M Na_2CO_3 to terminate the reaction. A TKO 100 fluorometer (Hoeffer, San Francisco, CA) calibrated by setting a 100-nM MU to 1,000 fluorescence units was used to measure fluorescence of the product, 4-methyl umbelliferone (4-MU). For each sample, results of GUS assay were plotted in a graph of OD_{405} (Y-axis) versus time in minutes and the GUS activity was expressed simply as the slope of the line. GUS activity from the maize endosperm suspension cells that had been bombarded with the naked gold particles (no DNA) was used as a control.

Using a luminometer (Monolight 1500; Analytical Luminescence Laboratory, San Diego, CA), luciferase activity was determined by measuring luminescence for 10

sec after mixing 20 μl of cell extract with 100 μl of luciferase assay reagent containing 20 mM tricine, pH 7.8, 1.07 mM $(MgCO_3)_4Mg(OH)_2 \cdot 5H_20$, 2.67 mM $MgSO_4$, 0.1 mM EDTA, 33.3 mM DTT, 270 μM coenzyme A, 470 μM luciferin and 530 μM ATP. LUC activity from the maize endosperm suspension cells that had been bombarded with the pLN (promoterless LUC plasmid) was used as a control. To correct differences in sample variability and transfection efficiency, the luciferase activity (in light units) was normalized with GUS activity, yielding the LUC/GUS ratio of each sample.

Results

Cloning and Characterization of the *Ae* Gene. After screening approximately 3 x 10^5 plaque-forming units of a genomic library prepared from maize seedlings (inbred B73), 12 lambda clones that strongly hybridized to the full-length *Ae* cDNA probe were isolated. These clones were hybridized to 5' or 3' cDNA fragment probes revealing that none of them contained both ends of the *Ae* gene. Based on restriction endonuclease maps, two clones, λ 3-2-1 and λ 7-2-1, were selected, subcloned, and sequenced. DNA sequences of the two clones revealed that the λ 3-2-1 clone containing the 5'-end of the *Ae* gene had approximately 1.5-kb of overlap with the λ 7-2-1 clone containing the 3'-end of the gene. A complete restriction map of the *Ae* gene was constructed by combining the two overlapping genomic clones which encompasses the entire coding region of the gene as well as large regions of 5'- and 3'- flanking sequences (Fig. 3).

Genomic Organization of the *Ae* Gene. Primer extension analysis was conducted to determine the

transcription initiation site of the *Ae* gene. A single major reverse transcription product was observed which co-migrated with a G residue in the sequencing ladder, indicating the transcription initiates mainly at a position which is located 28 bp downstream from a putative TATA box. The transcription initiation site was numbered +1 in the sequence shown in Figure 4. In addition to the TATA box, a number of potential *cis*-regulatory elements were found in the 5'-flanking region of the *Ae* gene (Fig. 4). The proximal promoter region from -300 to -87 contains two MRE boxes (TGCRCNC, R = purine, Y = pyrimidine), motifs essential for metal ion-dependent induction of both mouse and human metallothionein genes. In addition, this region contains four GC boxes (CCGCCC), sequences recognized by the mammalian transcription factor Sp1. In a region further upstream, sequences identical to a Hex (ACGTCA), a conserved element found in plant histone gene promoters; an I box (GATAGG), an element conserved in various RBCS genes encoding the small subunit of ribulose 1,5-bisphosphate carboxylase/oxygenase; and a RY repeat (ATGCCATG), a distal regulatory element which comprises a portion of the 28-bp legumin box were present at positions -999, -647, and -1491, respectively. Interestingly, DNA sequences from -505 to +463 are extremely high in G+C content (66.5%) and CpG dinucleotide frequency (10.1 per 100 bp) compared to an average of 40% G+C and 3.2 CpG per 100 bp for the rest of the genomic fragment. This region includes the proximal promoter region, two exons, and one intron. These characteristics are typical of CpG islands found in the mammalian genome, which are usually nonmethylated and flanked by methylated regions.

The genomic structure of the *Ae* gene was established by alignment with the published sequence of *Ae* cDNA. The transcribed region of the gene consists of 22 exons and 21 introns distributed over 16,914 bp in length. Figure 3 summarizes the organization of the *Ae* gene. The published cDNA sequence is identical to the corresponding genomic sequence except for three nucleotides present in exon 4 and exon 6. This may be due to the different genetic stocks used in the two studies. Table 2 shows the sequences around the exon/intron junctions and a list of putative branch point consensus sequences.

Table 2. List of introns and sequences of exon/intron borders in the *Ae* gene.

Intron number	Exon/	Putative intron branch point*	/Exon	Intron size(bp)	GC content (%)
1	ACTC	GTAA.....GTGAA..24[Y].GCAG	GGGG	106	51.9
2	GGAG	GTTC.....CTGAA..29..CCAG	GTAC	244	41.4
3	ATCG	GTAT.....TTCAA..21..ACAG	GTAC	1086	43.9
4	GCAG	GTAT.....ATAAC..24..GTAG	CGCG	76	25.0
5	ATTT	GTAT.....TTCAG..25..TTAG	TCTG	196	31.1
6	CAAA	GTAT.....CTAAA..22..GCAG	AATG	499	33.5
7	AAAG	GTAG.....TTAAC..23..ATAG	GTGA	81	40.7
8	AAGA	GTCT.....TTAAG..21..GCAG	GTAA	567	30.3
9	CCCG	GTAT.....ATAAT..23..TTAG	GAAC	774	32.0
10	TTGG	GTAA.....TTAAT..21..GCAG	ATAC	751	33.2
11	ATAG	GTAA.....TTAAC..22..GCAG	TCAT	4020	44.0
12	GGAA	GTAC.....TTTAT..49..GCAG	GTTT	86	34.9
13	ACAA	GTAA.....CTAAA..19..TTAG	GTAA	148	31.1
14	GTAA	GTGC.....TTCAA..20..TCAG	GTTA	3051	41.8
15	TCAA	GTAA.....TTCAA..19..ACAG	GCAA	872	31.8
16	CAAG	GTTA.....ATGAG..25..GCAG	GATA	457	32.6
17	CCTG	GTGA.....GTCAT..21..GCAG	AATG	144	30.6
18	CCTG	GTAA.....ATTAT..28..TCAG	GGTG	226	30.1
19	TGAA	GTAT.....ATGAA..19..GCAG	TTCA	266	31.2
20	TAAG	GTAT.....CTGAC..21..CCAG	GTGG	448	40.0
21	CGCC	GTAA.....CTAAC..24..GCAG	GACT	96	45.8

* Consensus sequences between introns are underlined.

[Y] Numbers indicate number of nucleotides between adjacent sequences.

The introns are relatively AT-rich (61%) compared to the exons (54%), and all have the conserved splice site sequences at their 5' and 3' ends, following the 'GT...AG' rule of plant introns. The introns vary in length from 76 bp (intron 4) to 4,020 bp (intron 11), and the exons vary in length from 43 bp to 303 bp. Exon 1 contains 100 bp of 5' untranslated DNA sequence, and exon 22 contains the translation stop codon (TGA) and 3' untranslated region. Although the canonical polyadenylation signal, AATAAA, was not found in the 3'-end of the gene, a similar sequence (AATTAAA) was observed 29 bp upstream of the polyadenylation site.

In addition, sequence analysis revealed that the 3'-flanking region of the *Ae* gene contains many direct repeat sequences and has a high degree of similarity to the pollen retroelement maize-2 (PREM-2), a copia-type retroelemt in maize which is expressed in a tissue-specific manner (Fig. 4). A 13-bp polypurine tract (AAAAAGGGGGAGA) which is present in PREM-2 and necessary for retroelement replication was also found upstream of the region which is very similar to the 3' long terminal repeat (LTR) of the PREM-2. Thus, the 3'-flanking region is likely to possess part of a PREM-2-type retroelement.

Genomic Southern Blot Analysis. Southern blot analyses were performed to determine the number of genes in the maize genome that are similar to *Ae*. When blots were probed with the full-length *Ae* cDNA under high-stringency conditions, at least two strongly hybridizing bands were observed in each lane. The band patterns agreed with the restriction map of the *Ae* genomic DNA, suggesting all the bands were derived from a single genetic locus. To confirm this, the blots were probed

with a small fragment of the *Ae* genomic DNA which does not have any restriction enzyme sites for BamHI, EcoRI, BglII and HindIII. As expected, only a single band was observed in every lane, supporting the conclusion that a single copy of *Ae* is present in the maize genome.

Analysis of the 5'-flanking Region of the *Ae* Gene. To identify the 5'-flanking regions necessary for *Ae* gene expression, we utilized a transient expression assay system in maize endosperm suspension cells. Iodine staining and northern blot analysis showed that the maize endosperm suspension cells actually produce starch and the genes involved in starch biosynthesis are expressed. In an initial experiment, a transcriptional chimeric construct containing the *Ae* gene fragment between -2964 and +100 linked to a luciferase (LUC) reporter gene in pUC119 was created and called pKL201. Since there are several examples of plant genes which are regulated by the first exon and/or intron sequences, a translational fusion construct (pKLN201) containing the corresponding region of the *Ae* gene was also created to test its effect on gene expression. pKLN201 was created by including an additional *Ae* DNA sequence from +101 to +329 in the plasmid pKL201. These plasmids were then tested by assaying LUC activity after introduction of DNA into maize endosperm suspension cells by particle bombardment. Plasmid pBI221 containing the CaMV 35S promoter linked to a GUS gene was used as an internal control to correct for transfection efficiency. The results showed that levels of LUC expression driven by the two constructs were almost the same , suggesting that the first exon and intron region of *Ae* is not necessary for high level gene expression in maize endosperm cells.

To define the promoter sequences important for

Ae gene expression in maize endosperm cells via transient expression assays, a series of 5' deletion mutants were derived from the plasmid pKL201 using available restriction enzyme sites and PCR techniques (Fig. 5A,B). The activity of each 5' deletion construct is presented in Fig. 5C. Two consecutive deletions of the *Ae* 5'-flanking sequence down to -1128 decreased the level of the LUC expression to approximately 40% of the full-length promoter level. However, the removal of an additional 353 bp, to -775, restored LUC activity (Fig. 5C). This suggests that potent positive and negative distal *cis*-regulatory elements may be located in the regions from -2964 to -1129 and from -1128 to -776, respectively. Although further deletions down to -160 did not significantly change the levels of LUC expression, a dramatic reduction in the promoter strength was observed when an additional 111 bp, to -49, was deleted. This indicates that very strong positive regulatory element(s) are located at the 111-bp region from -160 to -50. The presence of two GC boxes and one MRE motif in the region suggests the possibility that these conserved motifs may actually act as *cis*-regulatory elements essential for gene expression in maize endosperm cells.

EXAMPLE 3

Functional Analysis of *Sbe1* Promoter

We report in this example a functional analysis of the *Sbe1* promoter which reveals DNA sequence elements important for the high level and sugar responsive expression of the *Sbe1* gene in maize endosperm cells.

Materials and Methods:

Construction of Chimeric Plasmids. A transcriptional fusion of the *Sbe1* promoter to a luciferase (LUC) reporter gene was made as follows. A BamHI restriction enzyme site was first created just before the translation initiation site of the *Sbe1* gene by polymerase chain reaction (PCR): The DNA sequence between -253 and +27 of the *Sbe1* gene was PCR-amplified with PI-1 and PI-2 primers (Table 3).

Table 3. Oligonucleotides used in PCR to create *Sbe1* - LUC chimeric constructs.

Primer[a]	Sequence[b]	Annealing Region[c]
PI-1 U	CCAGCTCCACGGTTGTTCGTGT	-253 to -232
PI-2 L	cgatggatccTGTGACGGCGTGTGAGTCCC	+8 to +27
PI-3 L	agtcggatccTCAGGCGCACATTGCCGCCA	+209 to +228
PI-4 U	gactgagctcATACCAAATGAAGCCAGGAG	+5382 to +5401
PI-5 L	actggaattcGGAACAAGGAACGAAGAAAC	+5761 to +5780
PI-6 U	gatcaagcttACCAGCTCCACGGTTGTTCG	-254 to -235
PI-7 U	attcaagcttCAGATCCGGCTCAGGGTCAT	-196 to -177
PI-8 L	TGCGACAAGGAGGGGGCCAT	-165 to -146

[a] U and L indicate upper (sense) and lower (antisense) primers relative to the *Sbe1*, respectively.
[b] The lowercase letters designate restriction sites used for cloning.
[c] Numbers represent distance relative to the transcription start site (+1) of the *Sbe1*.

Four additional bases were included at 5'-end of the primers to provide for restriction enzyme sites at the ends of the PCR products for subsequent cloning. The bases were chosen randomly considering their effect on Tm value, dimer and stem-loop formation of the primers. —*Pfu* DNA polymerase (Stratagene, La Jolla, CA), which has proofreading activity, was used to enhance the fidelity of PCR amplification (*Pfu* DNA polymerase was used for all the following PCRs). Since an ApaI restriction enzyme

site (GGGCCC) is located immediately downstream of the 5' primer (PI-1) binding region of the *Sbe1* promoter, -203 to -198, the PCR product was digested with ApaI and BamHI. The resulting 236-bp fragment was then cloned into pBluescript SK$^-$ and sequenced to verify that no misincorporation had occurred in the DNA sequence during the PCR amplification (all the following PCR products were sequenced). Next, the 236-bp fragment was ligated to the 1991-bp SalI-ApaI *Sbe1* promoter fragment and cloned into plasmid pLN cut with SalI and BamHI (promoterless LUC-NOS gene in pUC119) (Montgomery et al., Proc. Natl. Acad. Sci. USA, 90: 5939-5943, 1993), thereby creating plasmid pKL101.

To construct a translational fusion of the *Sbe1* promoter containing the first exon and intron to a LUC reporter plasmid, the DNA sequence between -253 and +228 was amplified with the PI-1 primer and a 3' primer (PI-3) designed to anneal to the region just downstream of the first intron of the *Sbe1* gene. The 493-bp PCR product was digested with ApaI and BamHI, and the resulting 436-bp fragment was used to replace the ApaI-BamHI fragment in pKL101. This construct was called pKLN101. To make pKLM101 which contains the *Sbe1* promoter with four exons and introns, the 236-bp ApaI and BamHI fragment in pKL101 was replaced with the 1816-bp *Sbe1* genomic DNA fragment.

The plasmid pKLNS101 was derived from pKLN101 by replacing the nopaline synthase (NOS) 3' sequence with the native *Sbe1* 3'-flanking sequence. To accomplish this, two primers PI-4 and PI-5 were designed to amplify *Sbe1* DNA sequences containing the transcription stop signal and the polyadenylation site (from + 5382 to + 5780). A 419-bp PCR product was digested with SacI and EcoRI, and the resulting fragment was then used for

substituting a 255-bp SacI-EcoRI NOS 3' sequence in the pKLN101.

To create a series of 5' deletions in the *Sbe1* promoter, pKLN101 was first modified as follows: pKLN101 DNA was digested with HindIII and the resulting 7190-bp fragment lacking the 452-bp HindIII fragment was gel purified. The fragment was blunt ended by Klenow fill-in DNA synthesis and ligated with SalI linkers. After complete digestion with SalI, the DNA fragment was partially digested with BamHI to isolate the 1993-bp SalI-BamHI fragment, which was then gel purified and cloned into pLN cut with SalI and BamHI to produce pKLN101-1.

A series of 5' deletion mutants were made from the plasmid pKLN101-1 using the S1 nuclease based Erase-a-Base system (Promega, Madison, WI) to produce the 5' deletion series plasmids, pKLN102 to pKLN107. All constructs were sequenced with pUC/M13 reverse primer to verify deletion end points. For the -254 and -196 deletion constructs, two regions of the *Sbe1* promoter, -254 to -146 and -196 to -146, were PCR-amplified by primer PI-6 and PI-8, PI-7 and PI-8, respectively. Primers used in PCR to create the *Sbe1*-LUC constructs are shown in Table 2. Since each 5' primer, PI-7 and PI-8, contains a HindIII restriction enzyme site and a BstXI restriction enzyme site is located between -173 and -162, the PCR products were digested with HindIII and BstXI and the resulting fragments were used to replace the 2,047-bp HindIII-BstXI fragment of pKLN101.

Linker-Scanning Mutagenesis. A series of linker-scan mutations were introduced into the 60-bp DNA region from -314 to -255 as described by Kunkel et al (Kunkel et al., Methods Enzymol., 154: 367-382, 1987).

Briefly, the HindIII-BamHI (-314 to +235) fragment from pKLN105, containing the DNA region to be altered, was subcloned into the corresponding sites of a M13mp19 vector to produce a single-stranded template. In order to increase mutant recovery efficiencies, the template was prepared from an *E. coli dut⁻ ung⁻* strain (CJ236) which allows the incorporation of uracil into the newly synthesized DNA. Next, a set of oligonucleotides with 10-bp mismatches shown in Table 4 were annealed to the template and extended with T7 DNA polymerase. After addition of T4 DNA ligase, the resulting heteroduplexes were introduced into a wild-type E.coli strain (MV1190) to generate mutated double-stranded DNAs. DNA sequencing was performed to verify that the desired mutations were correctly introduced and no unintended mutations had occurred.

To create the mutated *Sbe 1* promoter-LUC constructs (pLS1-1 to pLS1-6), the HindIII-BamHI fragment in pKLN105 was replaced by each mutated DNA sequence.

Table 4. Oligonucleotides used in linker-scanning mutagenesis.

Constructs	Oligonucleotides*
pLS1-1	CCCGGTTTGCCTTTTTTgcTgcaggacAAGCTTGGCGTAATCAT
pLS1-2	TTGCACGCTTCCCGGTTgaCtgcaggcTATTTTATGTAAGCTTG
pLS1-3	GCCTTTGGGCTTGCACGtcTagatagcTGCCTTTTTTTATTTTA
pLS1-4	GGGCCGATTGGCCTTTGactgcaggtaCTTCCCGGTTTGCCTTT
pLS1-5	AGCTGGTTCTGGGCCGAcgactgcagaGGCTTGCACGCTTCCCG
pLS1-6	ACAACCGTGGAGCTGGTgtcGactatcTTGGCCTTTGGGCTTGC

* The mutated bases are shown in lowercase letters, and restriction sites used for convenience of screening are underlined.

Particle Bombardment. Suspension culture cells of maize (*Zea mays*) endosperm (inbred A636) provided by J. L. Anthony (DEKALB Genetics Corporation, Mystic, CT) were grown in 250-mL large-mouth Erlenmeyer flasks containing 80 mL of Murashige and Skoog basal salt medium supplemented with 0.4 mg/L thiamine 2 g/L asparagine and 30 g/L sucrose (Shannon and Liu, Physiol. Plant, 40: 285-291, 1977). The culture was maintained in the dark at 29°C on a rotary shaker (120 rpm) and was subcultured every 7 days by transferring a portion of the cell suspension into fresh medium.

For particle bombardment, about 600 mg (fresh weight) of actively growing cells 3 days after subculture was evenly distributed over the surface of a piece of filter paper (Whatmann #4, 55 mm in diameter) by vacuum filtration of 8 mL of suspension culture. The filter paper bearing the cells was then placed over three layers of filter paper (Whatmann #4, 70 mm in diameter) moistened with 5 mL of the liquid medium containing 12% sucrose and positioned in the middle of a 10 cm petri dish.

60 mg of gold microcarriers (1.6 μm particle size) were washed three times with 1 mL of 100% ethanol and twice with 1 mL of sterile deionized H_2O, resuspended in 1 mL of sterile deionized H_2O, and dispensed in 50-μL aliquots (3 mg/50 μL). The *Sbe1* Promoter-LUC constructs and a GUS reference plasmid (pBI221, Jefferson, Plant Mol. Biol. Rep., 5: 387-405, 1987) were co-precipitated onto the gold particles as follows: under continuous vortexing, the following were added in order to each 50-μL aliquot of gold particles: 5 μL of DNA (8 μg of LUC reporter plasmid, 4 μg of GUS reference plasmid), 50 μL of 2.5 M $CaCl_2$, and 20 μL of 0.1 M spermidine (free base,

tissue culture grade). The gold particles coated with DNA were pelleted in an Eppendorf centrifuge at 10,000 rpm for 10 sec, rinsed with 250 μL of 100% ethanol, and resuspended in 60 μL of 100% ethanol. Immediately after sonication, 8 μL of the DNA-coated gold particles were pipetted onto the center of macrocarriers (Bio-Rad, Hercules, CA) and dried in a low humidity environment.

A Bio-Rad PDS-1000/He Biolistic Particle Delivery system was used for particle bombardment. Bombardment parameters which were optimized include He pressure, gap distance (distance from power source to macroprojectile), and target distance (distance from microprojectile launch site to sample target). After optimization, all bombardments were performed in a dimly light room at 650 psi under a vacuum of 26 inches of Hg with a distance of 10 cm between the cells and the barrel of the particle gun. Following bombardment, the petri dishes were sealed with Parafilm and then incubated in the dark at 25°C for 24 hr.

GUS and LUC Assays. The bombarded cells were harvested from the plates by vacuum filtration, frozen in liquid nitrogen, and ground with a pestle and mortal to a fine powder. The powder was then transferred into a microfuge tube and extracted with cell culture lysis buffer containing 300 mM Tris-phosphate, pH 7.8, 2 mM DTT, 2 mM 1,2-diaminocyclohexane-N, N, N', N'-tetraacetic acid, 10% glycerol and 1% Triton X-100 (0.3 mL/g of tissue). Cell debris was pelleted in an Eppendorf centrifuge at 14,000 rpm for 10 min at 4°C and the supernatant was split into two aliquots for assays of GUS and LUC activity.

For fluorometric GUS assays (Jefferson, Plant Mol. Biol. Rep., 5: 387-405, 1987), 30 μL of the crude

extract was incubated at 37°C with 2 mM 4-methyl umbelliferyl glucuronide in 0.3 mL of GUS assay buffer (50 mM $NaPO_4$, pH 7.0, 10 mM EDTA, 0.1% Triton X-100, 0.1% Sarkosyl, 10 mM β-mercaptoethanol, 20% methanol). After 0, 1, and 2 hr of incubation, 0.1 mL aliquots were removed and added to 0.9 mL of 0.2 M Na_2CO_3 to terminate the reaction. A TKO 100 fluorometer (Hoeffer, San Francisco, CA) calibrated by setting a 100-nM MU to 1,000 fluorescence units was used to measure fluorescence of the product, 4-methyl umbelliferone (4-MU). For each sample, results of GUS assay were plotted in a graph of OD_{405} (Y-axis) versus time in minutes and the GUS activity was expressed simply as the slope of the line. GUS activity from the maize endosperm suspension cells that had been bombarded with the naked gold particles (no DNA) was used as a control.

Using a luminometer (Monolight 1500; Analytical Luminescence Laboratory, San Diego, CA), luciferase activity was determined by measuring luminescence for 10 sec after mixing 20 μL of cell extract with 100 μL of luciferase assay reagent containing 20 mM tricine, pH 7.8, 1.07 mM $(MgCO_3)_4Mg(OH)_2 \cdot 5H_20$, 2.67 mM $MgSO_4$, 0.1 mM EDTA, 33.3 mM DTT, 270 μM coenzyme A, 470 μM luciferin and 530 μM ATP. LUC activity from the maize endosperm suspension cells that had been bombarded with the pLN (promoterless LUC plasmid) was used as a control. To correct differences in sample variability and transfection efficiency, the luciferase activity (in light unit) was normalized with GUS activity, yielding the LUC/GUS ratio of each sample.

Nuclear Extract Preparation. Maize kernels (inbred B73) were harvested 30 days after pollination and frozen in liquid nitrogen. Nuclear extract was prepared

essentially according to the method described by Jensen et al. (Jensen et al., EMBO J., 7: 1265-1271, 1988). Protein concentration was determined using a BCA protein assay kit (Pierce, Rockford, IL), according to the manufacturer's instructions.

DNA Probe Preparation. The *Sbe1* promoter region from -314 to -255 was PCR-amplified with a forward primer (5'-GGACTTACATAAAATAAAAAAAGG CA) and a reverse primer (5'-TGCTAAGCTTTCTGGGCCGATTGGCCTTTG) which contain BamHI and HindIII restriction enzyme sites, respectively, at their 5' ends (underlined). The PCR product was digested with BamHI and HindIII, and the resulting fragment was cloned into pBlueScript SK$^-$ cut with BamHI and HindIII to create plasmid pRb4-1. The plasmid construct was verified with DNA sequencing. For electrophoretic mobility shift assays, the DNA fragment was cut out from the plasmid pRb4-1 with HindIII and BamHI, purified from agarose gels, and end-labeled with [α-^{32}P]-dCTP using the Klenow fragment.

Electrophoretic Mobility Shift Assay. The DNA-protein binding reaction was performed in 20 μL of solution containing 0.5 ng of labeled probe, 10 μg of nuclear protein, 1 μg of poly (dI-dC)·poly (dI-dC), 12% glycerol, 12 mM HEPES-NaOH (pH 7.9), 4 mM Tris-Cl (pH 7.9), 60 mM KCl, 1 mM EDTA and 1 mM DTT. After a 20-min incubation at room temperature, the samples were loaded into a 4% native polyacrylamide gel which was pre-run at 4°C for 1 hr at 150 V and electrophoresed for 2.5 hr at 150 V in Tris-glycine buffer at 4°C. Following electrophoresis, the gel was dried with a gel dryer (Bio-Rad) and exposed to Kodak X-ray film with two intensifying screens for 24 hr.

Northern Blot Analysis. Total RNA was isolated according to Vries et al. (Vries et al., Plant Mol. Biol. Manual, B6: 1-13, 1988) from maize endosperm suspension cells which had been incubated for 24 hr in the MS basal salt media supplemented with 0.4 mg/L thiamine, 2 g/L asparagine and various amounts of sucrose from 0% to 15%. Northern blot analysis was performed as described in Gao et al. (1996, *supra*). Radioactivity was detected with a PhosphorImager and quantified with the ImageQuant software program (Molecular Dynamics). To correct for minor loading errors between the lanes, the blot was wash at 95°C in a 0.1% (w/v) SDS solution to remove the ^{32}P-labeled *Sbe1* cDNA probe and rehybridized with a ^{32}P-labeled tomato cDNA for 26S rRNA.

Results:

Transcribed Regions of the *Sbe1* Gene are involved in gene expression. To determine whether the 5' flanking sequence of the *Sbe1* gene has all of the DNA elements necessary to initiate transcription, a 2217-bp fragment upstream of the translation start site (-2191 to +27) was fused to the luciferase (LUC) reporter gene in pUC119 (pKL101) as shown in Figure 6A. The chimeric plasmid was then introduced into maize endosperm cells via particle bombardment along with a reference plasmid containing the cauliflower mosaic virus (CaMV) 35S promoter linked to a GUS gene (pBI221, Jefferson, 1987, *supra*) to correct for transfection efficiency. However, only very low levels of LUC activity were detected relative to other promoters previously described.

Since many reports indicated that DNA sequences within transcribed regions such as exons, introns and 3' flanking regions are involved in the expression of genes

in either qualitative or quantitative manners, three different types of translational fusion constructs were created to test the effect of downstream elements on *Sbe1* gene expression. First, the 5'-flanking sequence as well as the first exon and intron of the *Sbe1* gene (-2190 to +228) were fused in-frame to the LUC reporter gene to make pKLN101 (Figure 6A, B). Second, the nopaline synthase (NOS) 3' sequence in the pKLN101 was replaced with the *Sbe1* 3' flanking sequence (399 bp in length), which contains the translation stop codon and polyadenlylation signal to create pKLS101. Finally, to determine whether an increase in the number of exon/introns enhances the gene expression, three more exons and introns from the *Sbe1* gene were added to the pKLN101 to make pKLM101 (-2190 to +1617).

The results of transient expression assays using the chimeric constructs are shown in Figure 6C. Inclusion of the DNA sequence from +28 to +228 containing the first exon and intron increased the level of LUC expression by 14-fold, suggesting that the first exon and intron region is required for high level expression of the *Sbe1* gene in maize endosperm cells. Since pKLN101 produce a fusion protein, however, we cannot completely rule out the possibility that the increase may be due to changes in enzyme activity and/or turnover rate caused by the added amino acid sequences. If the additional amino acids have a negative effect, the enhancement of LUC activity observed would be greater than 14 fold.

Replacement of the NOS 3' end in pKLN101 with the *Sbe1* 3' region did not have a significant effect on the level of LUC expression, implying that the *Sbe1* 3' untranslated region does not have indispensable control elements. However, it is still possible that the region

may be important for *Sbe1* gene expression in other cell types or inductive conditions.

Construct pKLM101 showed a slight reduction in LUC activity compared to pKLN101, indicating that additional exons and introns had an adverse effect on LUC expression in maize endosperm suspension cells. The adverse effect could be explained by inefficient splicing, resulting from the introduction of multiple copies of the plasmid into a single cell, or by formation of fusion protein consisting of the 5'-end of SBEI and luciferase, thus lowering LUC activity. Alternatively, it could also be actually due to the presence of negative *cis*-elements in this region.

5' deletion down to -314 did not significantly affect the *Sbe1* promoter activity. To identify promoter sequences critical for *Sbe1* expression in maize endosperm cells, a series of 5' deletion mutants were derived from pKLN101 as shown in Figure 7A. The activity of each 5' deletion construct is presented in Figure 7B. Removal of the sequences to -1332 caused a decrease in the level of the LUC expression, while deletion of an additional 422 bp, to -910, resulted in an increase in the activity of the construct. This suggests that potential positive and negative distal *cis*-regulatory elements may be located in the regions from -2190 to -1332 and from -1332 to -910, respectively. Further deletions down to -315 did not significantly affect the promoter activity, but a severe reduction in the activity was observed when an additional 169 bp, to -145, was deleted. The -72 deletion construct produced a level of the LUC activity slightly over background, showing that the minimal promoter is functional.

A 60-bp region is critical for the promoter activity. To further delimit sequences essential for high level expression of the promoter, two additional 5' deletions with about 60-bp intervals were created between -315 and -145. As shown in Figure 7, a deletion to -255 (pKLN108) severely reduced the expression of the LUC reporter gene, while a further deletion to -196 (pKLN109) did not further reduce promoter strength. This indicates that a strong positive regulatory element(s) is present in the 60-bp region between -315 and -255.

Linker-Scan Analysis reveals two *cis*-elements within the 60-bp region. Since the 5' deletion analyses indicated that the region of the *Sbe1* promoter from -314 to -255 is critical for the promoter activity, the 60-bp DNA fragment was further dissected by oligonucleotide-directed *in vitro* mutagenesis, as described by Kunkel et al. (1987, *supra*). A series of six different substitution mutants, designated pLS1 to pLS6, were created by altering the wild-type DNA sequence of the *Sbe1* promoter at 10-bp intervals. The mutations were made by creating transversion substitutions where possible, while at the same time introducing restriction enzyme sites for simplifying identification of the mutant forms.

The mutated constructs were tested for their promoter activity using the transient assay system, and the results of the experiments are shown in Figure 8. Mutations in the regions from -314 to -305 and -304 to -295, corresponding to pLS-1 and pLS-2, caused a decrease in the *Sbe1* promoter activity to 60% and 72% of wild-type (pKLN105) expression, respectively. The pLS3 construct showed almost the same level of the LUC expression as the wild-type promoter, suggesting the nucleotides from -294 to -285 are not important for the promoter activity in

maize endosperm cells. However, mutation in the pLS-4 region (-284 to -275) decreased promoter activity to 40% of the wild-type level. Also, other two mutants, pLS-5 and pLS-6, resulted in a reduction of the promoter activity to 55% and 50% of the wild-type promoter, respectively.

The 60-bp fragment interacts with DNA-binding proteins. In order to investigate the possibility that a nuclear protein(s) might interact with the 60-bp *Sbe1* promoter fragment from -314 to -255, electrophoretic mobility shift assays were performed. The 60-bp fragment was ^{32}P end-labeled with Klenow fill-in reaction and then incubated with nuclear extract prepared from 30 days after pollination (DAP) maize kernels (B73), which have been demonstrated to highly express *Sbe1* gene (Gao et al., 1996, *supra*). Two major shifted bands were observed in the lane containing nuclear extract (lane 2) compared to the control (lane 1). The bands were not detected after inclusion of proteinase K in the binding reaction (lane 7), indicating the shifted bands represent DNA-protein complexes.

Competition assays were conducted to determine whether or not the complexes are due to the binding of sequence-specific proteins. Inclusion of 10-fold and 100-fold excess of the unlabeled 60-bp fragment in the binding reaction significantly reduced formation of the complexes (lane 3 and 4), while the same amount of non-specific competitor DNA failed to compete for binding (lane 5 and 6). Thus, the complexes appear to be the results of sequence-specific interactions between a nuclear protein(s) and the DNA fragment, consistent with the functional identification of this region as an important regulatory element. Using six 60-bp fragments

of linker scan mutants (LS-1 to 6) as competitors, we have found that LS-1 did not affect the intensity of the lower band, although the rest of linker scan mutants abolished its formation. This suggests that the lower band may be the result of interactions between a trans-acting factor(s) and the sequence ACATAAAATA, located within LS-1. All of the linker scan mutants reduced the intensity of the slower migrating complex to varying degrees. Since LS-4, 5 and 6 were less effective competitors, wild type sequences spanning these regions (-284 to -255) may be involved in formation of this complex, however, binding may involve several overlapping regions in this fragment.

Expression of the *Sbe1* gene is sugar-regulated. The SBEs are expressed in a coordinate fashion with the granule-bound starch synthase (GBSS) and ADP-glucose pyrophosphorylase during maize endosperm development (Gao et al., 1996, *supra*). The ADP-glucose pyrophosphorylase gene (AGPase S) from potato and the genes encoding GBSS and SBE in cassava plants have been shown to be induced by an exogenous supply of sugars (Muller-Rober et al., Mol. Gen. Genet., 224: 136-146, 1990; Giroux et al., Plant Physiol., 106: 713-722, 1994; Salehuzzaman et al., Plant Sci., 98: 53-62, 1994). This led us to speculate that the *Sbe1* gene in maize may also be regulated by external sugar concentration.

To test this, maize endosperm suspension cells were incubated in MS media containing different concentrations of sucrose and their total endogenous RNAs were analyzed by northern blot hybridization. Sucrose was used in preference to other metabolizable sugars, because it is known to be the major sugar unloading from the pedicel tissue of maize kernels. Increase in sucrose

concentration from 0% to 9% elevated the *Sbe1* mRNA level by two-fold, and at higher concentrations the increase was reduced. Hexoses such as glucose, fructose and myo-inositol also increased the level of the transcript in a similar fashion. However, L-glucose and PEG 200 at concentrations calculated to have the same osmotic potential as a 9% sucrose solution (263 mM) did not exhibit any effect, indicating that the response is not an osmotic effect but a sugar-specific phenomenon. These results suggest that expression of the *Sbe1* gene in maize endosperm cells is regulated by sugar availability like other starch biosynthetic genes (Giroux et al., 1994, *supra*). This metabolic feedback mechanism may serve as a system to fine tune the expression levels of *Sbe* genes relative to physiological status of a plant. In fact, Shannon et al. (Shannon et al., Plant Physiol., 110: 835-843, 1996) showed that non-allelic starch mutants of maize accumulating high levels of sucrose in endosperm contain increased SBE activities compared to normal.

Since we determined that two *Sbe1* genes (*Sbe1a* and *Sbe1b*) with divergent 5'-flanking regions exist in the maize genome (Example 1), it was necessary to determine whether or not expression of the isolated *Sbe1* gene (*Sbe1a*) promoter is responding to external sucrose concentrations. To test this, a gene which is not regulated by sugar concentration was required for an internal control for the transient assay system. Since a CaMV 35S promoter has been used as a control in other studies investigating sucrose responsiveness of plant genes, the effect of sucrose on expression of the CaMV 35S promoter-GUS chimeric gene (pBI221) in maize endosperm cells was first investigated.

The plasmid pBI221 was bombarded into maize

endosperm suspension cells supplemented with 0% sucrose or 9% sucrose media and incubated at 25°C in the dark. After 48 hr incubation, GUS activity and protein concentration were measured from each sample to calculate specific GUS activity. The results showed that specific GUS activities of 9% sucrose samples were almost 2.5-fold higher than those of 0% sucrose samples, which is consistent with other reports. Since similar results were obtained from a ubiquitin promoter (pACH18) and -64 CaMV 35S minimal promoter which does not have an activation sequence (*as*)-1, a binding site for the transcription factor TGA-1a), it appeared that the elevated levels of expression by the CaMV 35S and ubiquitin promoters in 9% sucrose may be a general phenomenon simply due to an increase in energy source rather than a sugar-specific effect. Therefore, we reasoned that if the chimeric construct pKLN101 (the *Sbe1* promoter-LUC) is sugar-modulated, it will further enhance the level of LUC expression beyond the general increase at the higher sucrose concentration.

As shown in Figure 9, after normalization to GUS activity driven by the CaMV 35S promoter the plasmid pKLN101 still showed approximately two-fold greater LUC activity in 9% sucrose media than in 0% sucrose media. This is consistent with the result of the endogenous RNA analysis, indicating that the identified *Sbe1* gene is regulated by sugar availability. It also suggests that the nucleotide sequence containing a 2.2 kb 5'-flanking region and the first exon/intron of the *Sbe1* gene is sufficient to confer sugar responsiveness in maize endosperm cells.

Next, to delimit a region(s) necessary for the response, two deletion constructs, pKLN105 and pKLN106

were also tested in the transient expression system (Figure 9). Like pKLN101, pKLN105 (deletion end point -314) responded to a high sucrose concentration (9%) and increased LUC expression by approximately two-fold. However, pKLN106 (deletion end point -145) showed similar levels of LUC expression in both low and high sucrose conditions. These results suggests that the region between -314 and -145 contains a *cis*-regulatory element(s) necessary for the sugar response in maize endosperm cells. In addition, because the expression level was reduced in both + and - sucrose treated cells, other regulatory elements may also reside in this region.

Overexpression of mEmBP-1 protein represses the *Sbe1* gene expression. The canonical G-box sequence, CCACGTGG, was found in the 5'-flanking sequence of the maize *Sbe1* (-228 to -221) as well as the rice *Sbe1* gene (-170 to -163). This evolutionary conservation suggests a possible role of the G-box motif in the regulation of gene expression, although our 5' deletion analysis did not show it as an important regulatory element. It is known that the G-box motif resides in the promoters of many plant genes responding to a variety of different environmental and physiological stimuli and is often associated with additional regions which act as coupling elements determining signal response specificity.

To test whether or not the G-box in the maize *Sbe1* promoter is interacting with a G-box binding protein in maize, mEmBP-1, which is a homologue of the wheat EmBP-1 (Guiltinan et al., Science, 250: 267-271, 1990) and expressed during endosperm development, EMSA and DNase I footprint analyses were performed. As expected, the analyses clearly showed that EmBP-1 interacts with the G-box sequence *in vitro*. Since EmBP-1, a member of

the basic leucine zipper (bZIP) transcription factors, is implicated in ABA-induced *Em* gene expression in wheat. The data prompted us to ask two questions. First, is the *Sbe1* gene expression regulated by ABA concentration? Second, can mEmBP-1 protein transactivate *Sbe1* gene expression? Transient expression assays failed to show a relationship between exogenous ABA concentrations (1 to 100 μM) and the *Sbe1* promoter activity in the maize endosperm suspension cells, suggesting the G-box in the *Sbe1* promoter is not ABA-responsive.

To address the second question, a chimeric construct containing the cauliflower mosic virus (CaMV) 35S promoter fused to the full-length mEMBP-1 cDNA (35S-mEmBP-1) was created and co-introduced with the plasmid pKLN101(a full-length *Sbe1* promoter-LUC) into the maize endosperm suspension cells. We predicted overexpression of mEmBP-1 protein would enhance the LUC expression driven by the *Sbe1* promoter, since mEmBP-1 is known as a bZIP transcription activator. Contrary to the prediction, overexpression of mEmBP-1 protein actually resulted in a significant reduction (5-fold) of the *Sbe1* promoter activity as shown in Figure 10. The effect was apparently selective for the *Sbe1* promoter, since mEmBP-1 had little effect on expression of a LUC reporter gene linked to the ubiquitin promoter (pACH18). Interestingly, substitution of the G-box sequence (CCACGTGG) in pKLN105 with TTGAACTA did not cause a reduction in promoter activity, suggesting that the G-box sequence is not required for high level expression of the *Sbe1* gene in maize endospem cells.

Discussion

The expression pattern of the maize *Sbe1* gene has been investigated in almost all maize tissues (Gao et al., 1996). The *Sbe1* gene is constitutively expressed at a low level in vegetative tissues, while it is modulated during kernel development. Especially in the endosperm, *Sbe1* mRNA began to accumulate to high levels at the onset of rapid starch deposition. These findings suggest that the expression of *Sbe1* is regulated by certain factors which vary in concentration or activity during kernel development.

As a step toward understanding regulatory mechanisms controlling *Sbe1* gene expression, we analyzed the *Sbe1* promoter regions using a transient gene expression system. Transient expression assays showed that expression driven by the maize *Sbe1* promoter greatly depends on the presence of the DNA region spanning the first exon and intron of the maize *Sbe1*. Addition of the DNA sequence (+28 to +228) containing the first exon and intron of the *Sbe1* gene into the transcriptional chimeric construct (pKL101) increased reporter gene expression in maize endosperm suspension cells up to 14-fold. Since such DNA sequences containing transcriptional stimulating effects are useful in investigations of gene expression in plant cells and for plant genetic engineering, it will be necessary to determine whether or not the DNA sequence has the ability to increase gene expression under the control of other promoters. There are several examples of plant genes which are regulated by DNA sequences within the transcribed region. Among them, the first exon and intron sequences of the maize *Sh1* gene are one of the best examples studied so far (Vasil et al., Plant Physiol., 91: 1575-1579, 1985; Maas et al., Plant Mol.

Biol., 16: 199-207, 1991; Clancy et al., Plant Sci., 98: 151-161, 1994). The *Sh1* exon appears to have two separate *cis*-elements which act independently to increase gene expression via different mechanisms. One of the elements may contain a novel promoter element which has the ability to interact with transcription factors binding upstream. The other acts possibly at the level of translation efficiency or mRNA stability. The enhancing effect of the *Sh1* intron is likely the result of an increase in the level of mature cytoplasmatic mRNA level like the maize *Adh1* first intron.

5' deletion analysis of the maize *Sbe1* promoter revealed several *cis*-regulatory elements affecting the promoter activity in maize endosperm cells. Of special interest was the identification of the 60-bp positive element located in the region from -314 to -255 relative to the transcription initiation site. Further investigation of the region using linker-scan analysis identified at least two separate regions, -314 to -295 and -284 to -255, which are critical for gene expression in maize endosperm cells.

Interestingly, the -314/-295 region has striking similarity with the sucrose-responsive element (SURE-1) of the potato patatin-1 promoter (Grierson et al., Plant J., 5: 815-826, 1994), which has been shown to interact with a sucrose-inducible nuclear protein(s). Grierson et al. demonstrated that a 100-bp patatin-1 promoter fragment encompassing SURE-1 is sufficient to confer sucrose responsiveness. DNA sequences similar to the -314/-295 region are also found in the promoter regions of other sugar-inducible genes, such as maize sucrose synthase (Shaw et al., Plant Physiol., 106: 1659-1665, 1994), *Arabidopsis* β-amylase (Mita et al., Plant

Physiol., 107: 895-904, 1995), and potato sporamin (Ohta et al., Mol. Gen. Genet., 225: 360-378, 1991). This finding along with the sugar enhanced expression of the *Sbe1* demonstrated by northern blot analysis and transient expression assay strongly suggest that the conserved sequences may be implicated in mediating sugar responsiveness of the *Sbe1* gene. This is further supported by our recent finding showing that the -314/-196 region of the *Sbe1* promoter is sufficient to confer the sucrose responsiveness to -64 CaMV 35S minimal promoter. Since high sucrose concentration media were used for the transient expression assays to maximize gene expression, it is understandable that mutation of this region decreased the level of LUC expression. It remains to be tested whether or not other *Sbe* genes are also sugar-modulated. To date, we have not detected sugar-dependent DNA binding activity associated with the *Sbe1* promoter.

In potato and cassava plants, sugars have been shown to regulate expression of genes involved in starch biosynthesis. Our results demonstrated that the maize *Sbe1* is also modulated by sugar concentration. Such sugar effect was not due to change in osmotic potential, because L-glucose and PEG which are osmotically active did not affect *Sbe1* gene expression. Recently, it has been reported that hexokinase is involved in sensing sugar concentration in higher plants, and sugar signaling mediated through hexokinase is uncoupled from sugar metabolism.

Sequence comparison between the rice (Kawasaki et al., 1993, *supra*) and maize *Sbe1* genomic DNAs (Example 1) revealed that the 5'-flanking sequences proximal to the protein-coding regions are highly divergent except

for the canonical G-box sequences (CCACTGG) which are located in similar positions relative to the corresponding transcription initiation sites. This evolutionary conservation between the species led us to postulate that the G-box may be involved in regulation of the *Sbe1* gene expression possibly in response to one of the environmental or physiological stimuli, even though we failed to show the importance of the G-box in the *Sbe1* promoter activity using the 5' deletion analysis. It is possible that G-box dependent mechanism controlling *Sbe1* promoter activity could not be appraised in our endosperm suspension cultured cells. This hypothesis is supported by the results showing interaction of the G-box with mEmBP-1 protein *in vitro* and repression of the *Sbe1* promoter activity by over-expression of mEmBP-1. Along with these, the finding that disruption of the G-box sequence (CCACGTGG) in pKLN105 did not cause a reduction in promoter activity led us to speculate that the G-box and its binding proteins are involved in down-regulation of the *Sbe1* gene expression rather than up-regulation. Although a specific role for the G-box motif in *Sbe1* gene expression has not been identified, there is a possibility that the G-box in the *Sbe1* promoter may play a critical role under different environmental conditions or in different tissues.

It has been noted that mutations decreasing starch accumulation in maize endosperm also reduce storage protein synthesis, implying possible interactions between these pathways. It has been shown that mutations affecting synthetic events in one biosynthetic pathway affect expression of genes in both pathway, and demonstrated that expression of genes involved in starch and storage protein synthesis of the maize endosperm are

coordinately regulated. Elevation in sugar concentration or alteration of the osmotic potential of the endosperm was proposed to be a possible candidate for the primary signal triggering this coordinate expression. In this context, knowledge of the *Sbe* promoter elements and their associated regulatory proteins may eventually lead to a better understanding of the regulatory mechanisms controlling all of the starch biosynthetic genes and the genes encoding storage proteins in maize endosperm.

While certain of the preferred embodiments of the present invention have been described and specifically exemplified above, it is not intended that the invention be limited to such embodiments. Various modifications may be made thereto without departing from the scope and spirit of the present invention, as set forth in the following claims.

We claim:

1. An isolated nucleic acid molecule for controlling expression of genes in transformed plant cells, which comprises a segment of an *Sbe1* gene from a plant species selected from the group consisting of *Zea spp.*, *Tripsacum spp.* and *Sorghum spp.*, the segment commencing at a location about 3,000 bases upstream from a transcription initiation site of the gene, and ending at a location about 250 bases downstream from the transcription initiation site.

2. The nucleic acid molecule of claim 1, wherein the plant species is *Zea mays*.

3. The nucleic acid molecule of claim 1, isolated from a gene having a coding sequence at least 60% homologous with the coding sequence defined by the exons of SEQ ID NO:1.

4. A fragment of the nucleic acid molecule of claim 1, comprising a segment commencing at about 3,000 bases upstream from the transcription initiation site and terminating about 25 bases downstream from the transcription initiation site.

5. A fragment of the nucleic acid molecule of claim 1, comprising a segment located between about 25 and 250 bases downstream from the transcription initiation site, the fragment being capable of increasing promoter activity of homologous or heterologous promoters.

6. A fragment of the nucleic acid molecule of claim 1, comprising a segment located between about 145 - 315 bases upstream from the transcription initiation site, the fragment comprising elements that render expression of the gene sugar-regulatable.

7. An isolated nucleic acid molecule for regulating expression of genes in transformed plant cells, which comprises a segment of a gene encoding a *Sbe1* gene from a plant species selected from the group consisting of *Zea spp., Tripsacum spp.* and *Sorghum spp.*, the segment comprising a 3' untranslated region commencing at a stop codon for gene's coding sequence, and ending at a location about 5900 bases downstream from the gene's transcription initiation site.

8. The nucleic acid molecule of claim 6, wherein the plant species is *Zea mays*.

9. The nucleic acid molecule of claim 6, isolated from a gene having a coding sequence at least 60% homologous with the coding sequence defined by the exons of SEQ ID NO:1.

10. A DNA segment for effecting expression of coding sequences operably linked to the segment, isolated from a gene whose coding region hybridizes under stringent conditions with a coding region defined by exons of SEQ ID NO:1, the segment comprising a promoter and a transcription initiation site.

11. The DNA segment of claim 10, which further comprises an element included in a first exon of SEQ ID

NO:1, the element being capable of increasing promoter activity of homologous or heterologous promoters operably linked thereto.

12. The DNA segment of claim 10, which further comprises an element that confers sugar-regulatability on expression of the coding sequences.

13. The DNA segment of claim 10, isolated from a maize *Sbe1* gene.

14. A DNA segment for modulating expression of coding sequences operably linked to the segment, isolated from a gene whose coding region hybridizes under stringent conditions with a coding region defined by exons of SEQ ID NO:1, the segment comprising a polyadenylation signal.

15. The DNA segment of claim 14, isolated from a maize *Sbe1* gene.

16. An isolated nucleic acid molecule for controlling expression of genes in transformed plant cells, which comprises a segment of a *Ae* gene from a plant, the segment commencing at a location about 3,000 bases upstream from a transcription initiation site of the gene, and ending at a location about 100 bases downstream from the transcription initiation site.

17. The nucleic acid molecule of claim 16, isolated from a *Ae* gene of a monocotylednous plant.

18. The nucleic acid molecule of claim 17, isolated from a maize *Ae* gene.

19. The nucleic acid molecule of claim 16, isolated from a gene having a coding sequence at least 60% homologous with the coding sequence defined by the exons of SEQ ID NO:2.

20. A fragment of the nucleic acid molecule of claim 16, comprising a segment located between about 50 and and 160 bases upstream from the transcription initiation site.

21. An isolated nucleic acid molecule for regulating expression of genes in transformed plant cells, which comprises a segment of a gene encoding a *Ae* gene from a plant, the segment comprising a 3' untranslated region commencing at a stop codon for gene's coding sequence, and ending at a location about 20,500 bases downstream from the gene's transcription initiation site.

22. The nucleic acid molecule of claim 21, isolated from a *Ae* gene of a monocotyledenous plant.

23. The nucleic acid molecule of claim 22, isolated from a maize *Ae* gene.

24. The nucleic acid molecule of claim 21, isolated from a gene having a coding sequence at least 60% homologous with the coding sequence defined by the exons of SEQ ID NO:2.

25. A DNA segment for effecting expression of coding sequences operably linked to the segment, isolated from a gene whose coding region hybridizes under stringent conditions with a coding region defined by exons of SEQ ID NO:2, the segment comprising a promoter and a transcription initiation site.

26. The DNA segment of claim 25, isolated from a maize *Ae* gene.

27. A DNA segment for modulating expression of coding sequences operably linked to the segment, isolated from a gene whose coding region hybridizes under stringent conditions with a coding region defined by exons of SEQ ID NO:2, the segment comprising a polyadenylation signal.

28. The DNA segment of claim 27, isolated from a maize *Ae* gene.

29. A chimeric gene comprising a coding sequence operably linked to one or more DNA segments selected from the group consisting of:

a) an expression regulatory element, isolated from a gene whose coding region hybrizes under stringent conditions with a coding region defined by exons of SEQ ID NO:1, the element comprising a promoter, a transcription initiation site and, optionally:

i) an element included in a first exon of SEQ ID NO:1, capable of increasing promoter activity of homologous or heterologous promoters operably linked thereto; and

ii) an element that confers sugar-

regulatability on expression of the coding sequence;

b) a expression regulatory element, isolated from a gene whose coding region hybrizes under stringent conditions with a coding region defined by exons of SEQ ID NO:2, the element comprising a promoter and a transcription initiation site;

c) an expression modulatory element, isolated from a gene whose coding region hybrizes under stringent conditions with a coding region defined by exons of SEQ ID NO:1, the element comprising a 3' untranslated region; and

d) an expression modulatory element, isolated from a gene whose coding region hybrizes under stringent conditions with a coding region defined by exons of SEQ ID NO:2, the element comprising a 3' untranslated region.

30. The chimeric gene of claim 29, inserted into a vector for transforming a cell.

31. A cell transformed with the vector of claim 30.

32. The transformed cell of claim 31, which is a plant cell.

33. The transformed cell of claim 32, which is a maize plant cell.

34. A transgenic plant produced by regenerating the transformed plant cell of claim 32.

35. A reproductive unit of the transgenic plant of claim 34.

37. A promoter isolated from a maize *Sbe1* gene.

38. A promoter isolated from a plant *Ae* gene.

39. The promoter of claim 38, isolated from a *Ae* gene of a monocotyledenous plant.

40. The promoter of claim 39, isolated from a maize *Ae* gene.

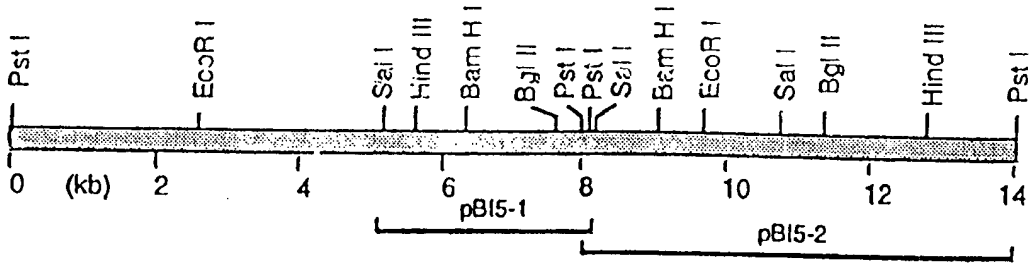

Pst I
EcoR I
Sal I
Hind III
Bam H I
Bgl II
Pst I
Pst I
Sal I
Bam H I
EcoR I
Sal I
Bgl II
Hind III
Pst I
0
(kb)
2
4
6
8
10
12
14
pBI5-1
pBI5-2

Fig. 1A

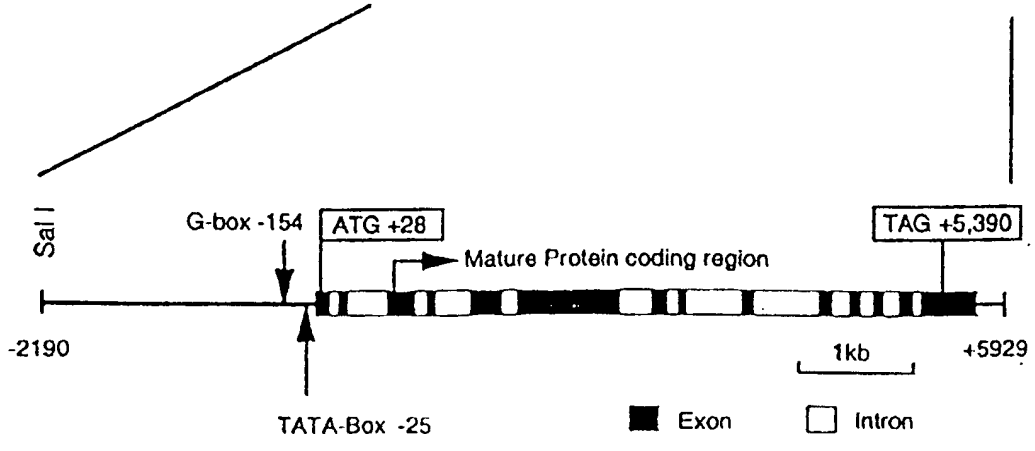

Sal I
G-box -154
ATG +28
Mature Protein coding region
TAG +5,390
-2190
1kb
+5929
TATA-Box -25
Exon
Intron

Fig. 1B

Fig. 2

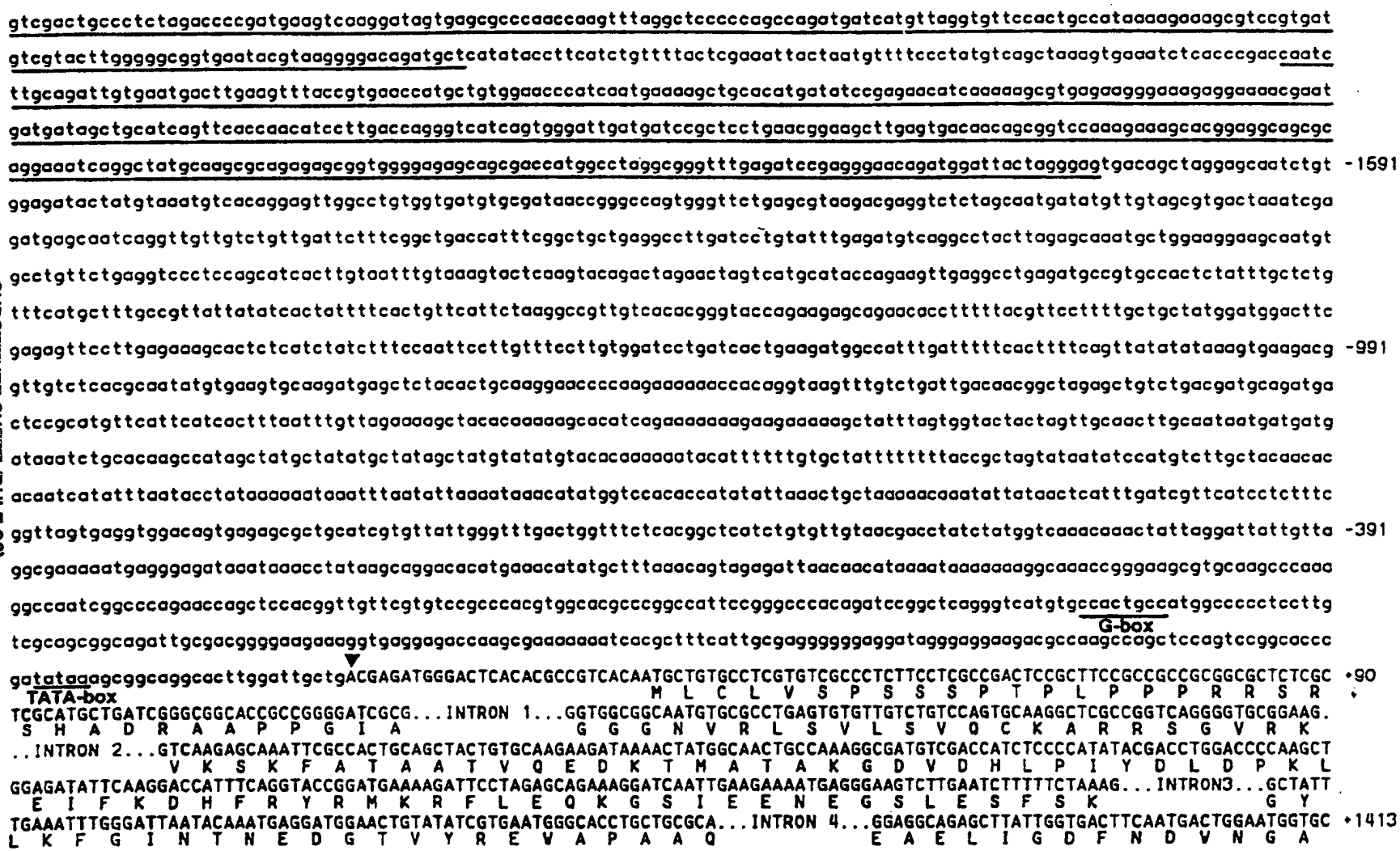
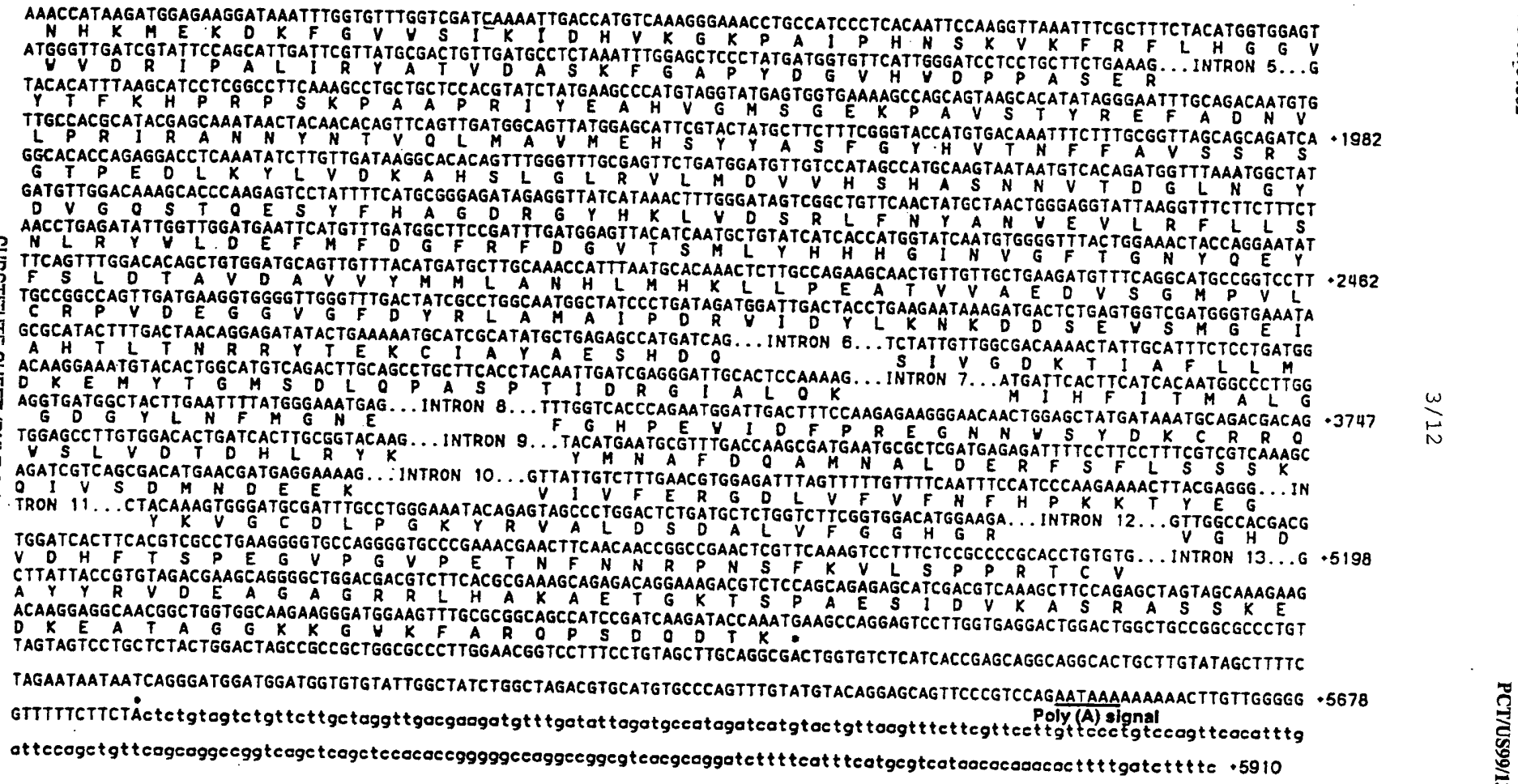


Fig. 2, cont.



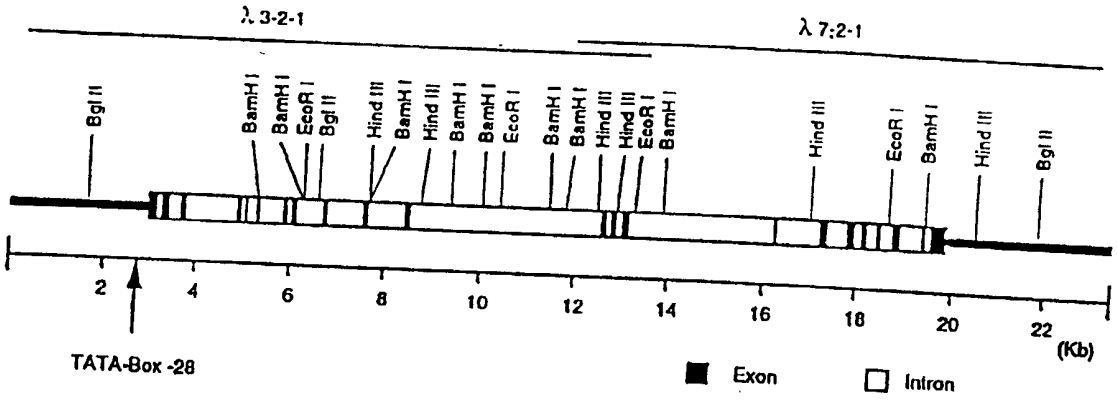
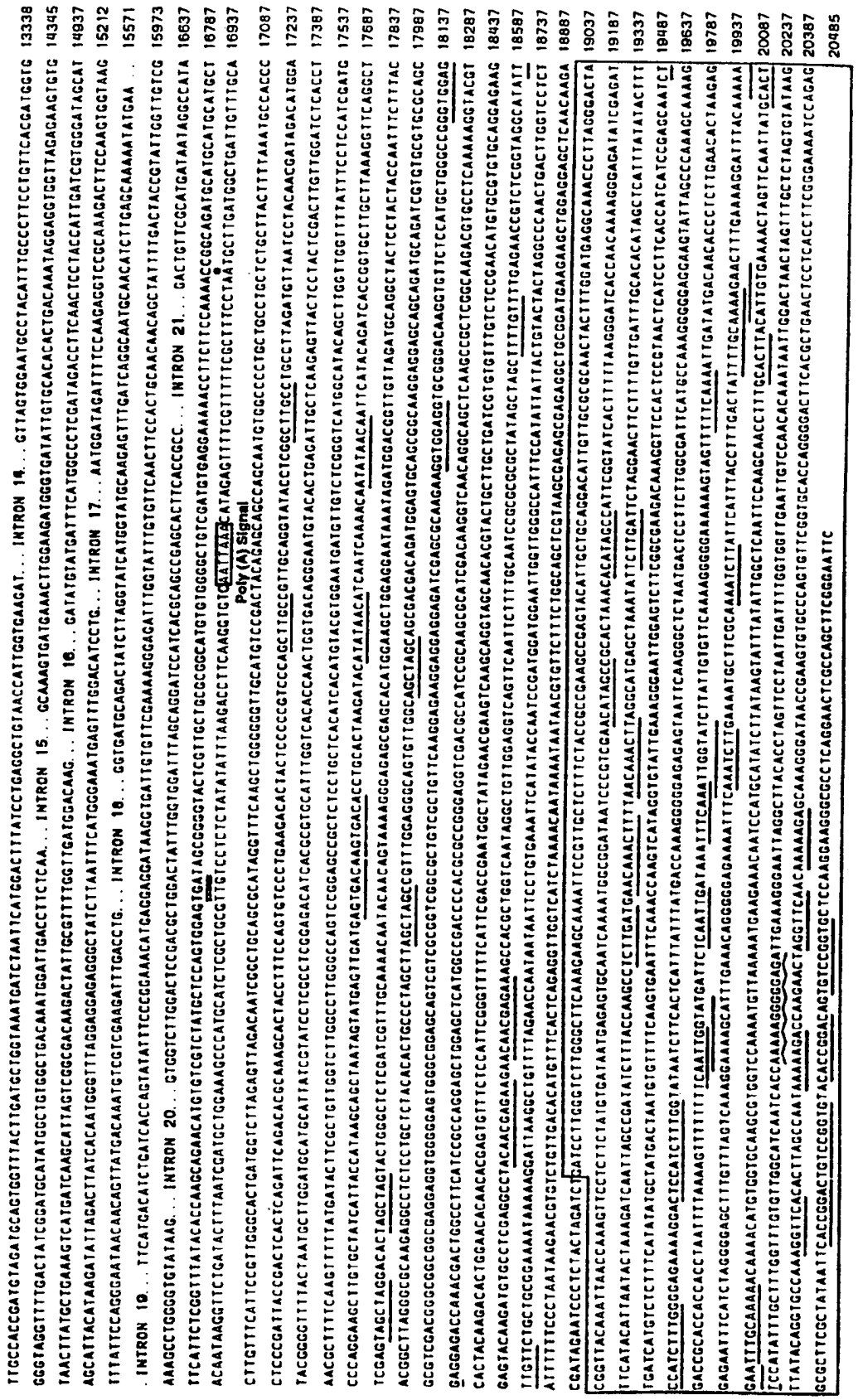

λ 3-2-1
λ 7-2-1
Bgl II
BamH I
BamH I
EcoR I
Bgl II
Hind III
BamH I
Hind III
BamH I
BamH I
EcoR I
BamH I
BamH I
Hind III
Hind III
EcoR I
BamH I
Hind III
EcoR I
BamH I
Hind III
Bgl II
2
4
6
8
10
12
14
16
18
20
22
(Kb)
TATA-Box -28
Exon
Intron

Fig. 3

Fig. 4

Fig. 4, cont



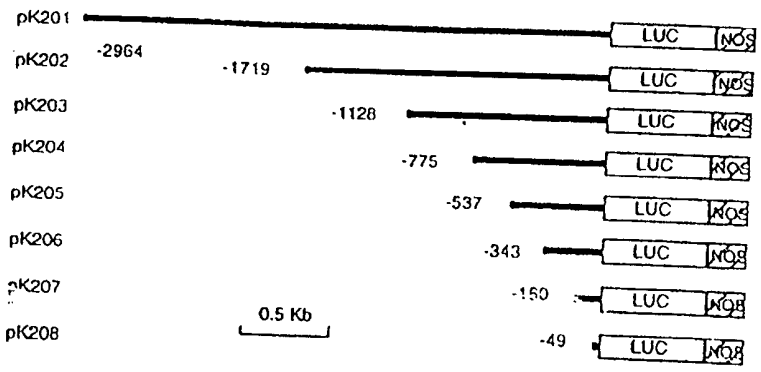


Fig. 5A

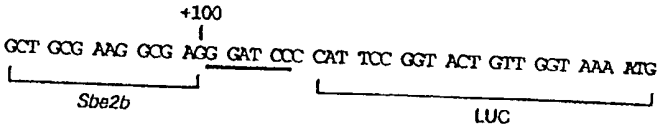


Fig. 5B

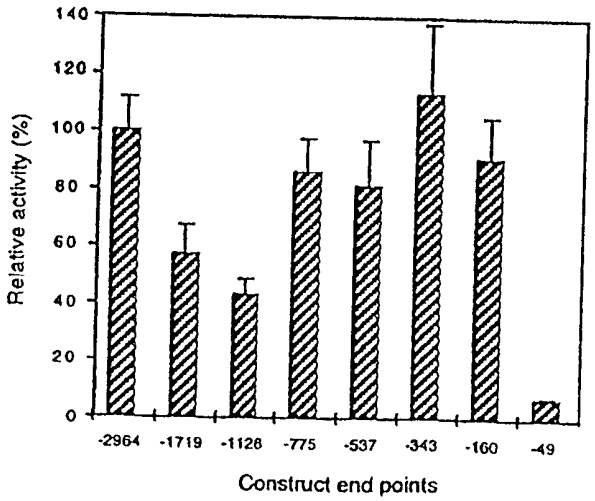


Fig. 5C

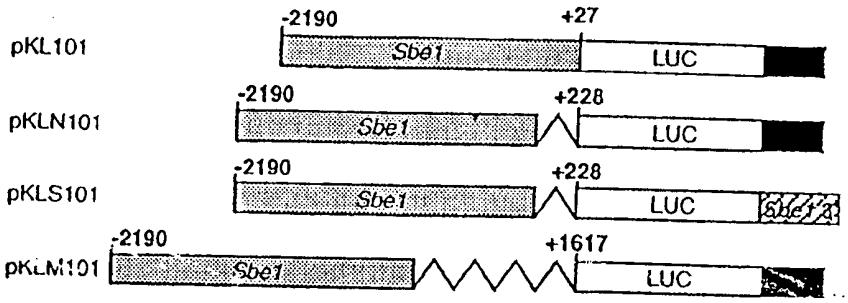


Fig. 6A

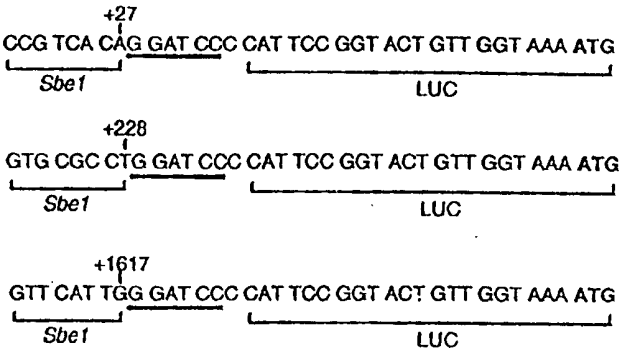


Fig. 6B

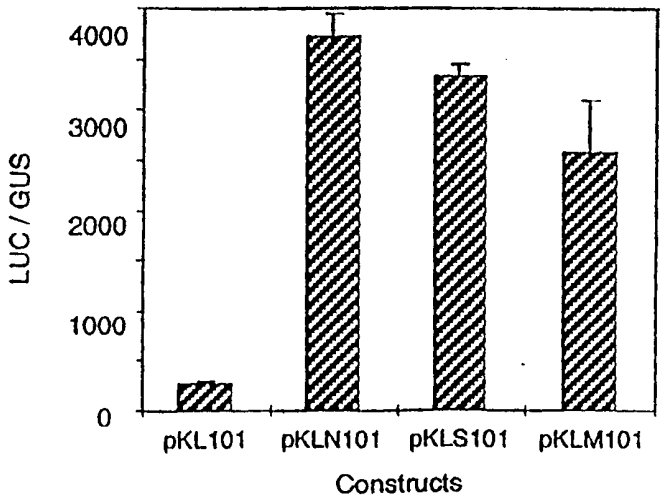


Fig. 6C

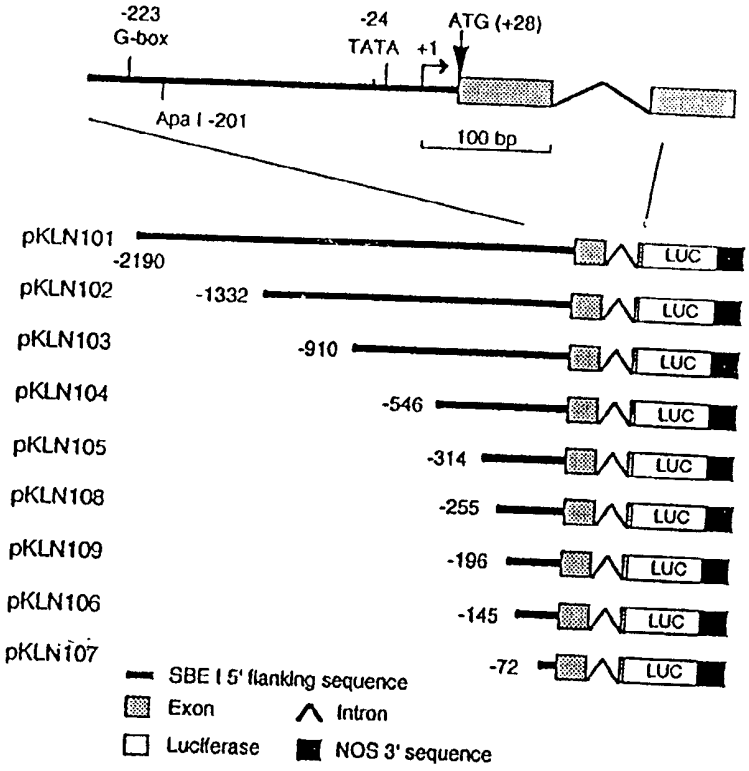

-223
G-box
-24
TATA
+1
ATG (+28)
Apa I -201
100 bp
pKLN101
-2190
pKLN102
-1332
pKLN103
-910
pKLN104
-546
pKLN105
-314
pKLN108
-255
pKLN109
-196
pKLN106
-145
pKLN107
-72
LUC
SBE I 5' flanking sequence
Exon
Intron
Luciferase
NOS 3' sequence

Fig. 7A

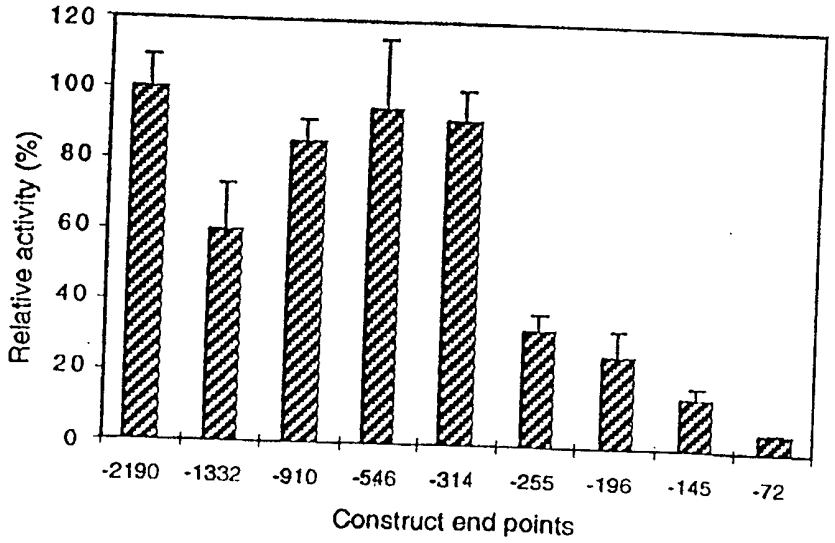

120
100
80
60
40
20
0
Relative activity (%)
-2190
-1332
-910
-546
-314
-255
-196
-145
-72
Construct end points

Fig. 7B

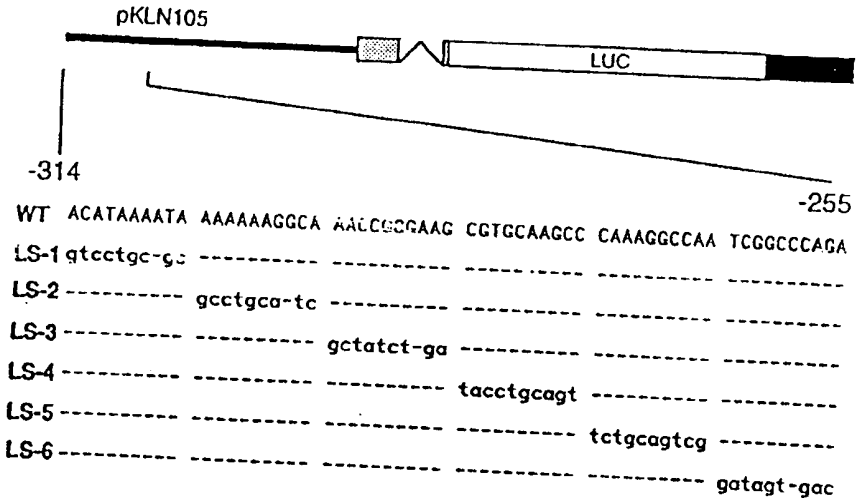

pKLN105
LUC
-314
-255
WT ACATAAAATA AAAAAAGGCA AACCGCGAAG CGTGCAAGCC CAAAGGCCAA TCGGCCCAGA
LS-1 gtcctgc-gc ---------- ---------- ---------- ---------- ----------
LS-2 ---------- gcctgca-tc ---------- ---------- ---------- ----------
LS-3 ---------- ---------- gctatct-ga ---------- ---------- ----------
LS-4 ---------- ---------- ---------- tacctgcagt ---------- ----------
LS-5 ---------- ---------- ---------- ---------- tctgcagtcg ----------
LS-6 ---------- ---------- ---------- ---------- ---------- gatagt-gac

Fig. 8A

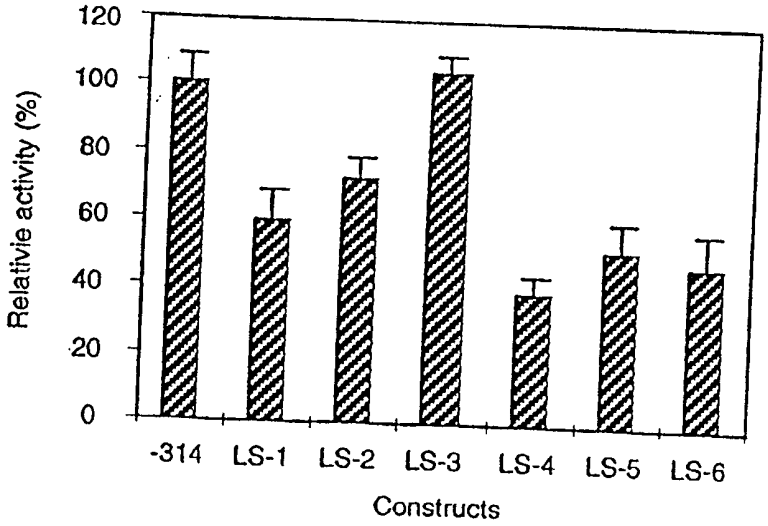

120
100
80
60
40
20
0
Relativie activity (%)
-314
LS-1
LS-2
LS-3
LS-4
LS-5
LS-6
Constructs

Fig. 8B

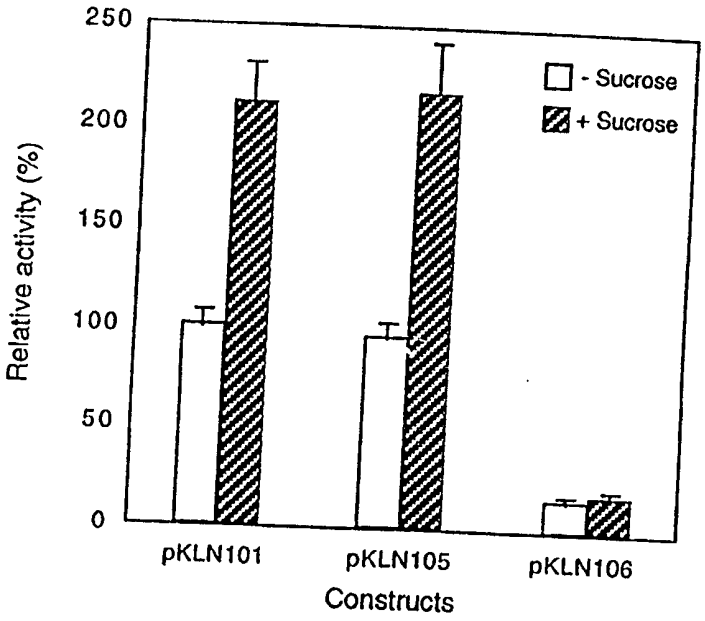

250
200
150
100
50
0
Relative activity (%)
- Sucrose
+ Sucrose
pKLN101
pKLN105
pKLN106
Constructs

Fig. 9

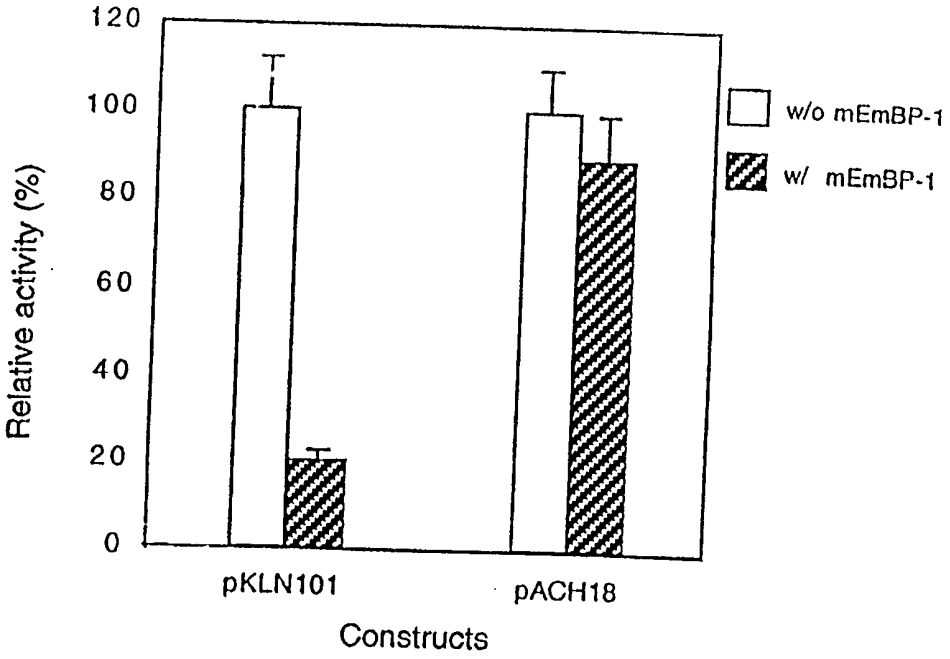

120
100
80
60
40
20
0
Relative activity (%)
pKLN101
pACH18
Constructs
w/o mEmBP-1
w/ mEmBP-1

Fig. 10

SEQUENCE LISTING

<110> Mark A. Guiltinan
Kyung-Nam Kim

<120> Expression Control Elements from Genes
Encoding Starch Branching Enzymes

<130> PennState 1465

<150> US 60/089,049
<151> 1998-06-12

<150> US 60/089,050
<151> 1998-06-12

<160> 18

<170> FastSEQ for Windows Version 3.0

<210> 1
<211> 8119
<212> DNA
<213> Zea mays

<400> 1

gtcgactgcc ctctagaccc cgatgaagtc aaggatagtg agcgcccaac caagtttagg 60
ctcccccagc cagatgatca tgttaggtgt tccactgcca taaaagaaag cgtccgtgat 120
gtcgtacttg ggggcggtga atacgtaagg ggacagatgc tcatatacct tcatctgttt 180
tactcgaaat tactaatgtt ttccctatgt cagctaaagt gaaatctcac ccgaccaatc 240
ttgcagattg tgaatgactt gaagtttacc gtgaaccatg ctgtggaacc catcaatgaa 300
aagctgcaca tgatatccga gaacatcaaa aagcgtgaga agggaaagag gaaaacgaat 360
gatgatagct gcatcagttc accaacatcc ttgaccaggg tcatcagtgg gattgatgat 420
ccgctcctga acggaagctt gagtgacaac agcggtccaa agaaagcacg gaggcagcgc 480
aggaaatcag gctatgcaag cgcagagagc ggtggggaga gcagcgacca tggcctaggc 540
gggtttgaga tccgagggaa cagatggatt actagggagt gacagctagg agcaatctgt 600
ggagatacta tgtaaatgtc acaggagttg gcctgtggtg atgtgcgata accgggccag 660
tgggttctga gcgtaagacg aggtctctag caatgatatg ttgtagcgtg actaaatcga 720
gatgagcaat caggttgttg tctgttgatt ctttcggctg accatttcgg ctgctgaggc 780
cttgatcctg tatttgagat gtcaggccta cttagagcaa atgctggaag gaagcaatgt 840
gcctgttctg aggtccctcc agcatcactt gtaatttgta aagtactcaa gtacagacta 900
gaactagtca tgcataccag aagttgaggc ctgagatgcc gtgccactct atttgctctg 960
tttcatgctt tgccgttatt atatcactat tttcactgtt cattctaagg ccgttgtcac 1020
acgggtacca gaagagcaga acacctttt acgttccttt tgctgctatg gatggacttc 1080
gagagttcct tgagaaagca ctctcatcta tctttccaat tccttgtttc cttgtggatc 1140
ctgatcactg aagatggcca tttgattttt cacttttcag ttatatataa agtgaagacg 1200
gttgtctcac gcaatatgtg aagtgcaaga tgagctctac actgcaagga accccaagaa 1260
aaaaccacag gtaagtttgt ctgattgaca acggctagag ctgtctgacg atgcagatga 1320
ctccgcatgt tcattcatca ctttaatttg ttagaaaagc tacacaaaaa gcacatcaga 1380
aaaaaagaag aaaaagctat ttagtggtac tactagttgc aacttgcaat aatgatgatg 1440
ataaatctgc acaagccata gctatgctat atgctatagc tatgtatatg tacacaaaaa 1500
atacattttt tgtgctattt tttttaccgc tagtataata tccatgtctt gctacaacac 1560
acaatcatat ttaataccta taaaaaataa atttaatatt aaaataaaca tatggtccac 1620
accatatatt aaactgctaa aaacaaatat tataactcat ttgatcgttc atcctctttc 1680
ggttagtgag gtggacagtg agagcgctgc atcgtgttat tgggtttgac tggtttctca 1740
cggctcatct gtgttgtaac gacctatcta tggtcaaaca aactattagg attattgtta 1800

ggcgaaaaat gagggagata aataaaccta taagcaggac acatgaaaca tatgctttaa 1860
acagtagaga ttaacaacat aaaataaaaa aaggcaaacc gggaagcgtg caagcccaaa 1920
ggccaatcgg cccagaacca gctccacggt tgttcgtgtc cgcccacgtg gcacgcccgg 1980
ccattccggg cccacagatc cggctcaggg tcatgtgcca ctgccatggc cccctccttg 2040
tcgcagcggc agattgcgac ggggaagaaa ggtgaggaga ccaagcgaaa aaaatcacgc 2100
tttcattgcg agggggggagg atagggagga agacgccaag ccagctccag tccggcaccc 2160
gatataaagc ggcaggcact tggattgctg acgagatggg actcacacgc cgtcacaatg 2220
ctgtgcctcg tgtcgccctc ttcctcgccg actccgcttc cgccgccgcg gcgctctcgc 2280
tcgcatgctg atcgggcggc accgccgggg atcgcggtaa gctgcggcga tccgggggct 2340
ggtggattta ctacttttgc ttttttttct tctgtactga tggttcatag tttcgcaggg 2400
tggcggcaat gtgcgcctga gtgtgttgtc tgtccagtgc aaggctcgcc ggtcaggggt 2460
gcggaaggta gatctttccc cagattagaa cacgttgtta cggacaaaat ggaatttggt 2520
agtagttttg gagagagaaa aaaaactctg gaatttgtgg attgggtcct gctcgctttg 2580
gatttggctc tagtgctttg ttatcagtag taatcatttt accaagtcta gcggatctca 2640
gatttcaact gaagttgttt aaatttggt ggttggtttt cttcagccat cctaatcatt 2700
tattacataa taattagttg aaataaagtg ctatatttga tatttctgtg gttgctgcag 2760
cagctaatta gcttcactac attttcacgt gttagcttct attgctcgtg taattctgaa 2820
tctggcttta acgcactgtg caaggtcaag agcaaattcg ccactgcagc tactgtgcaa 2880
gaagataaaa ctatggcaac tgccaaaggc gatgtcgacc atctccccat atacgacctg 2940
gaccccaagc tggagatatt caaggaccat ttcaggtacc ggatgaaaag attcctagag 3000
cagaaaggat caattgaaga aaatgaggga agtcttgaat ctttttctaa aggttaggct 3060
taatacattc aaatgctaat acaaggacca tttcacaata tttatagcac ctgcccatgc 3120
cgcaaaatat tatctgcatc atcaacattt ccggattttg tatttttttc aggctatttg 3180
aaatttggga ttaatacaaa tgaggatgga actgtatatc gtgaatgggc acctgctgcg 3240
cagtaagttc taatgttgtc atgcaaacat gatgtactgg cggggtatcg ttttttccca 3300
ttttgcttgt gaagagatat cgtctgatag taactatatt aaaaaaaaag aacagctaaa 3360
acacaggaga gtaatcaatg tttgggagta gaaaactgtt cggatatttc attcttgtgc 3420
aagactgtca tgtcttagct gaggaggcta ttttttttct tgctgtgatt acataatgct 3480
gttttcttct tatgacaata tactttccac tattttaata gattgatgca tttgacttga 3540
gttttttact catgggtgta gggaggcaga gcttattggt gacttcaatg actggaatgg 3600
tgcaaaccat aagatggaga aggataaatt tggtgtttgg tcgatcaaaa ttgaccatgt 3660
caaagggaaa cctgccatcc ctcacaattc caaggttaaa tttcgctttc tacatggtgg 3720
agtatgggtt gatcgtattc cagcattgat tcgttatgcg actgttgatg cctctaaatt 3780
tggagctccc tatgatggtg ttcattggga tcctcctgct tctgaaaggt ctctttctac 3840
tgtcctggac cacattgtaa acttaagcaa tagaccttag cagacaaaaa tatatgacaa 3900
aattatagta ccttttacct ctggtttgca ggtacacatt taagcatcct cggccttcaa 3960
agcctgctgc tccacgtatc tatgaagccc atgtaggtat gagtggtgaa aagccagcag 4020
taagcacata tagggaattt gcagacaatg tgttgccacg catacgagca aataactaca 4080
acacagttca gttgatggca gttatggagc attcgtacta tgcttctttc gggtaccatg 4140
tgacaaattt ctttgcggtt agcagcagat caggcacacc agaggacctc aaatatcttg 4200
ttgataaggc acacagtttg ggtttgcgag ttctgatgga tgttgtccat agccatgcaa 4260
gtaataatgt cacagatggt ttaaatggct atgatgttgg acaaagcacc caagagtcct 4320
attttcatgc gggagataga ggttatcata aactttggga tagtcggctg ttcaactatg 4380
ctaactggga ggtattaagg tttcttcttt ctaacctgag atattggttg gatgaattca 4440
tgtttgatgg cttccgattt gatggagtta catcaatgct gtatcatcac catggtatca 4500
atgtggggtt tactggaaac taccaggaat atttcagttt ggacacagct gtggatgcag 4560
ttgtttacat gatgcttgca aaccatttaa tgcacaaact cttgccagaa gcaactgttg 4620
ttgctgaaga tgtttcaggc atgccggtcc tttgccggcc agttgatgaa ggtggggttg 4680
ggtttgacta tcgcctggca atggctatcc ctgatagatg gattgactac ctgaagaata 4740
aagatgactc tgagtggtcg atgggtgaaa tagcgcatac tttgactaac aggagatata 4800
ctgaaaaatg catcgcatat gctgagagcc atgatcaggt accctgcatt atataataac 4860
tttataagga agtatagttg gatgcagcaa agtattttat ttccttgtgg ggtagacaac 4920
acatggggca agtatctttt aatcagacag tactatattt catcacattt ctatcctgat 4980
tgctagtaac tggattagtt attcataaaa aattaggccc tcctaagcat gtttaagaat 5040
ttgcaatcac atatgacctg tttatcttat actggaaatt aaaatcatgc aaatttttcag 5100
tctattgttg gcgacaaaac tattgcattt ctcctgatgg acaaggaaat gtacactggc 5160
atgtcagact tgcagcctgc ttcacctaca attgatcgag ggattgcact ccaaaaggtt 5220
ccattctcct caagttctgg ttgaacctgg ttttcaataa ttcttgaatg agttccctca 5280
atataccttg cttattgtgt tctgtacact ccagatgatt cacttcatca caatggccct 5340

tggaggtgat ggctacttga attttatggg aaatgaggtg aaatctcggt cttttgaaaa 5400
atgtttccca ttaataagcc tgcataaacc ttttgtattt ttttttctatt tgaggtttca 5460
gcatttagag tcgacccctct attttctcag taactgaagt tcctaatggc tatttgaatg 5520
gttaatgtta attccgagta gcgctcgttg cacatgaaaa tgtggtctgt ttatggctgc 5580
taatcctttt gtgacttgtc caatatactc gtacaacatg ggatacatgt aacccttcac 5640
tatcaagcat cgcttaaatg tcaacatcat acagagatat tatcaagtct ttctcctgca 5700
ctatcctgga atgaaaaaaa agagttgttt catatcgtat ctgtactttt acaaatgagt 5760
gtgatctttg gagaagttag ttgacggaat ggtccactgt attttgtttg ttatatgttt 5820
ctccttattc atagtgtatt caattgatgt gcaattctct tctagtttgg tcacccagaa 5880
tggattgact ttccaagaga agggaacaac tggagctatg ataaatgcag acgacagtgg 5940
agccttgtgg acactgatca cttgcggtac aaggttatgt ctatgaatgc aatccttata 6000
agatttttgt tctggcacca ttctaggctc tctttgacct ttacctcttc catttatcgc 6060
tgtctcctat actagctaga tctgtctctg agactagcta gatgtcagtt ccttactact 6120
attacactct tgagttgagc tgggtccaga tatgcttttc cacactttgt tttgctgccc 6180
cactgcccta ctatagtaaa ttgcaccaac cccgtacatg ttactcaata aacaggttgg 6240
attggcctgt gcaaacctag tattttgtaa gttgacctgg ctttatgtat tgccattgtt 6300
gaataaaaat aattttgatg cattgattgg tatgtcattt tcataacttt gcagacctag 6360
gaatgtactt ctcaatcatt ggttgttttg ttgatagaaa ttattctgtt tacagcatat 6420
ggtggactaa atatccaagt gaacgtttgt tttgtctgtt ttttttttcta cccctcaagt 6480
ttacaattta tcccggctaa catattaatg ttttctcctt ggctctaatg ctgaacagta 6540
catgaatgcg tttgaccaag cgatgaatgc gctcgatgag agattttcct tcctttcgtc 6600
gtcaaagcag atcgtcagcg acatgaacga tgaggaaaag gtaaggattt cagaatactt 6660
ggtatgagag tgttgcgacg tttctttctt tcttaggaag agtggtaact ctagcttgtg 6720
tgctgtattt tgccgattaa ggcgagcttc aatttaggag tggatggttt gtcagcatta 6780
ttttttctac ttcgtttcag gttattgtct ttgaacgtgg agatttagtt tttgttttca 6840
atttccatcc caagaaaact tacgaggggt aagtcaccag ttgtaaaacc ctgtcttttc 6900
agagtccctt ttgcctgtgg tataatataa tatagtgtgc ttacctccca tctctgctct 6960
gcaacttata tgcagctaca aagtgggatg cgatttgcct gggaaataca gagtagccct 7020
ggactctgat gctctggtct tcggtggaca tggaagagta agtagtgacg gtagacgctg 7080
aaagatgttt tttttttta atttttgccc tactcccgta gttggggccg gccggccgtc 7140
gagaatcgat catatcaggc tgtcgtgttt gcgagactca tggatgctgt gacgcaggtt 7200
ggccacgacg tggatcactt cacgtcgcct gaaggggtgc caggggtgcc cgaaacgaac 7260
ttcaacaacc ggccgaactc gttcaaagtc ctttctccgc cccgcacctg tgtggtaatg 7320
ttcacttact ctcagctgaa ctgatgagct agaatgtatc cgccctgaca aaccgtcctg 7380
ctgccaggct tattaccgtg tagacgaagc aggggctgga cgacgtcttc acgcgaaagc 7440
agagacagga aagacgtctc cagcagagag catcgacgtc aaagcttcca gagctagtag 7500
caaagaagac aaggaggcaa cggctggtgg caagaaggga tggaagtttg cgcggcagcc 7560
atccgatcaa gataccaaat gaagccagga gtccttggtg aggactggac tggctgccgg 7620
cgccctgtta gtagtcctgc tctactggac tagccgccgc tggcgcccct ggaacggtcc 7680
tttcctgtag cttgcaggcg actggtgtct catcaccgag caggcaggca ctgcttgtat 7740
agcttttcta gaataataat cagggatgga tggatggtgt gtattggcta tctggctaga 7800
cgtgcatgtg cccagtttgt atgtacagga gcagttcccg tccagaataa aaaaaaactt 7860
gttggggggt ttttcttcta ctctgtagtc tgttcttgct aggttgacga agatgtttga 7920
tattagatgc catagatcat gtactgttaa gtttcttcgt tccttgttcc ctgtccagtt 7980
cacatttgat tccagctgtt cagcaggccg gtcagctcag ctccacaccg ggggccaggc 8040
cggcntcacg caggatcttt tcatttcatg cgtcataaca caaacacttt tgatcttttc 8100
agacaaaaaa gatggatcg 8119

<210> 2
<211> 23449
<212> DNA
<213> Zea mays

<400> 2
ctcgagaatt atttttgata attcatcgga ccgtcattgt tcatggtcag tggcacacaa 60
gtccggaccg cctccaaagg atagtctttg gcgtgcgcgg gtgagccaac tgttagctga 120
cgtggcaccc aatccagagg tgcactagac taatccggtg cccgtcagcg aaggaaagtg 180
tcaaccagaa tcctggactc atgacgcatg aaccgatcca gtacccccgc aagtaagctg 240
atatttgaca tctttttgca agaaaggagc aatgactatg gggtcttttg gagctataaa 300

agcaccccca tggcgacctc cttcagaacc caagcactcc aagagtgtca catcactccg 360
actctcttct gtaactcatt ctagtgattt agtgagatca atgcgtcatt tctgagctgt 420
tcttgtgtat gtgtgacttt gcgcttttgt gatcttctct cttgagtttt gctgctacgt 480
tttggtcttg tatttgtaat ctctcccctc tcttgctttg agttgtaact ctgatattgt 540
gtacgactgc aagagactct aaattatgga gattccttgc aaacagggat ataagtgata 600
aggaaggacg tggcactcaa tttcgatctt tggatcactt gagagcggct gattgcaatc 660
cttgtcgttg gcaaccacaa cgtgaagtag tcaagcattt taggcttgac cgaaccacga 720
gaaaaatcat cgtgtcttgt gtgcttatcc acttgtgatt tttctccctc tttctaagtc 780
tctcagattc acttgtaata ttactcttaa tatattatat atcttgaaag agcaaacaag 840
ttgaagatga cttactttct tcttctctct atttaaactt ggcttggttt cactaacttc 900
catttccaga ccaaacttgt gtttttagtg ttgtttttgc agggtcacct attcaccccc 960
tctaggtgct ctcaagaggt gttagctatg tttccttagg cgtgtggcgg tcacctttтc 1020
gatagatatg aataagttcg aactttcttt agaaacaata agtcgcaaat tcaacaaagt 1080
actaaatgcg ctatacaaca tgtctagtga cgtaattaag ctaaaggacg accaatttcg 1140
tgagattcac cctagactat tggaggtgcg tttttggcaa cattgtaggg attgtataga 1200
actatcgatg gaagccactt tcctgtaatc gtacctaggc ggcatgtata cgcgtcgcaa 1260
aatgtcatgg cagtatgcga cttcgatatg tggtttacct ttgccgtgac tggttggttg 1320
ggctcagttc atgatctgta ggtactacaa gagactctca ttgccacgac aacatctttc 1380
cccatccacc tgaaggtaat gttttgggcg aacaaaaata attgtcacaa atgttataaa 1440
tttgaatagc attgctaatg aacttgtttt ggtaatgcat gaatatatta cttggtcgat 1500
tccggtcagc gtaacaggtt gggctacttt gccccgtata agggacaaaa gtatcacgtt 1560
acagaacgac aacatgggag acatcccgta gatgagaaag gggtattcaa ctatgcacat 1620
tcatccctaa gaaatgtcat agagcggtca tttgatgtgt ttgaaggtga agtggcagat 1680
cttgcttagt ctcccatgtt tctcggtgcg gaagcaataa aaaataatta taacttgtat 1740
gaccctacat aactttatta gggatagtgc tatacacaac cacgacttgg aaattttgtg 1800
tctgggaagc atgtaactga tgtaggggta acagaaacta gtagtggacc attaggtgag 1860
ttagacacga gtgagtttcc agatgacatt tatgctgcgc tagtgacatg acctaccatt 1920
tgtatttcat agatgtggca tgaacatgtt tatgttgtaa tgttctacgt cagtgcatta 1980
tgtaattgtg taattaattt aaggttgccc tagtagcaat gaaatacaaa tctccatcgg 2040
gatcggaggg aattaaggtg aaaatgaact tattttctct ccaacccctt caatctcgaa 2100
ggggatttga gtttccaaac tagatcctaa aagcaagtat aataaggtga tgtaggcgag 2160
ctgtaggata taacacatca gatttgtgat gatatgaaag aaaaaaaatg aagagagaat 2220
gaaagcgaac tgttgctcac ggactctaag atataatata agaggaagag atgagctaag 2280
tattaaatgt ataaatatat ttttagttag ttatatatga taggacaata tcataaaact 2340
cactttgtgc catatcgtta aacttgctgt agcttcaccg ttaatcgata aaaaatatta 2400
aaaaagcatc tcccgtttgc tcgactgctc gagtgtgcaa aaaaaaaggc acgacggacc 2460
cgcgcgctga cgcggtgagc cgcaagtccg caacggcgcg gccgcgcgct aggaaaaaag 2520
cctcgcccgt gaagcgaact ccctctcctt cgaccttcgt tcttccactg cggcctgcgc 2580
gacccgtgca gctgcgcgct ccacctggcc gcgctggggc ccacaccgcc tggcatctgg 2640
agcattgccc ccggacttcg cgcggccgcc cgcagccccg ctccccaccg aaaagcgaag 2700
cgattgccat ccccacgcca ccgcgaagca caaggtcccg ccctgcacga tcagcaggac 2760
ctcgccacgc cgccgctgga gctgcgcgtg cgcgtgtgcg cttggaccga cgcgcaacgg 2820
cctgcctcga ccgcccgtgc acgccactgc tcatgcagcc gtccgcctcg cccccgcccc 2880
gaactgccga ggtcgcgtga acgcccactc ccctcaccgc tcgtctccgt gctatatagg 2940
cagcccgcgc ccctcctaat tgtagccctg cagtcaccca gagcagaccc ggatttcgct 3000
cttgcggtcg ctggggtttt agcattggct gatcagttcg atccgatccg gctgcgaagg 3060
cgagatggcg ttccgggttt ctggggcggt gctcggtggg gccgtaaggg ctccccgact 3120
caccggcggc ggggagggta gtctagtctt ccggcacacc ggcctcttct taactcgtaa 3180
tgatcctgca actcctcctc cctctctgat caagtgtggg cctgattcgg gtctgtatgc 3240
gagtgttgtg gtggtgaact ggtgaattgg tgatgcatgc agggggtgct cgagttggat 3300
gttcggggac gcacggggcc atgcgcgcgg cggccgcggc caggaaagcg gtcatggttc 3360
ctgagggcga gaatgatggc ctcgcatcaa gggctgactc ggctcaattc cagtcggatg 3420
aactggaggt tcgtcatcca ctcgtcactt tcatgcattt tatcacataa ttcacctgaa 3480
agtctacatc tacatgcatt tctgattttt acctcttttt ggatgctatt tgagaacaat 3540
gagacacacg attagtgaga tgcccaaacg gttaaacgtt ccatggagct ccacaagtct 3600
gtactgtacc accatttaga gttgtccata caatcgtctt ctgaacattt cgtctcttcg 3660
gcacggttcc aggtaccaga catttctgaa gagacaacgt gcggtgctgg tgtggctgat 3720
gctcaagcct tgaacagagt tcgagtggtc cccccaccaa gcgatggaca aaaaatattc 3780
cägattgacc ccatgttgca aggctataag taccatcttg agtatcggta tgtattactt 3840

gcttctactg caccactact gaaacatttc aagacgttcc cgtctgtcat gatatatata 3900
agggtgtgtt tggttgggga gtgagggaga ataaagtggt tccattctaa ttttctggac 3960
gtttagtttc caacaaaata agagtggagc ggctcctgga tctccatatg gaaatttact 4020
ataagtagtt ggaacgctcc cgctccacaa aaacgacgga tgcgagcgct ctctcggacg 4080
tgagcgctct cctggctctc ctcccctcta cactcatacg gctctcaaac caaacaaaga 4140
atggaatgac tccgctctat tctcgtgaca agattctcgt gttatcttaa ttcttgcata 4200
cttttaggtg cgtttggttg gggagccact gggatggagc ggctccattc cagtttctaa 4260
aaattgagcc atttcattct gtgtttgaca agcagaacgg aaccacttca ttttttgttt 4320
ggttgaagag tagaatagag cggagccgct ccatttattg tttggttgga gagccatatg 4380
actgttgcgt ggaggagagg gatgagagcg cttgcgtccg cgagagcgct cgcatcggat 4440
cgttttggtg gagcgagagc atcccaaata tttgctattt atagactcca ggaaccgctc 4500
cgctctcctc catggtaacc aaacatcaaa aaaataagaa tggagccgtt ccattcccct 4560
ccgctcttca accaaacaca ttagtcaggt tcttcatagt gttgctggaa ggtcatatga 4620
atggcttagt caggttcgtt ttcacttgaa atctttcata ctgctttgat gcctaaagta 4680
ttccttatgg tttcaccaag cattacaatt gaacggacaa gatatgttct ctttcagagg 4740
atatttacag aaatgttagc cagacaatat agaggcaaca gcgggttcta attctttggc 4800
aaccttcatt tccatccctt ctgtattctt cttataagga ttcctatata taacccatgc 4860
gaacaccact atccatctca ctgtgctttc aacttccttg aataaaagaa caggtacagc 4920
ctctatagaa gaatccgttc agacattgat gaacatgaag gaggcttgga agccttctcc 4980
cgtagttatg agaagtttgg atttaatcgc aggtattctt taacatgaag tgtttatctt 5040
ttatcctata tactagcaaa taactttgta tcttattttc caatgtagcg cggaaggtat 5100
cacatatcga gaatgggctc ctggagcatt tgtatgatct ttcttcaatt attctaatct 5160
tattgcttat catgatacaa tactagttcc atgtttcatt atgagaatga tcactctccc 5220
agctacgaag ctactcatat ttaataaatc ttttaccttc aaaatatata cattgtcaga 5280
cttagctcct agtgttattc agaattgact cctgctttct atcttagtct gcagcattgg 5340
tgggtgactt caacaactgg gatccaaatg cagatcgtat gagcaaagta tgcttatgcc 5400
ttcagtaaac attatatatt tcacttatta gattggtttg tttatctttt tcattaactc 5460
aatatatgtg cactctattt aatcaacttt ccaaccggaa agtgcatatg ttgcattcac 5520
tggagtcagt tcctacaatt gactctgcat ataagtgtgc tattagatac ttgtgttgct 5580
gctgatactc ttagaagtca atgtttccaa actgcacatg tgatacctgt attctaatgt 5640
ctcgatatct ttacaaattg catgttccta taataggatg aattgaatac tcatacttag 5700
tgtgcatgga cttggagtag gtctaattct taatgattct cgaagagttg tttgtttgta 5760
agggttctta ccccttttatt tcttcttaat ataataatga tacacaactc tcttgtatgt 5820
tagcgaaaaa aaaagaaaac ttactgatcc tcgcatagtc taaatcgagg tactttgttc 5880
ttgcagaatg agtttggtgt ttgggaaatt tttctgccta acaatgcaga tggtacatca 5940
cctattcctc atggatctcg tgtaaaggta gccgctttac ctcattggtt gttgtttttt 6000
gctggggagc atctccaaaa ttaactggtg tgttttgacc atatataggt gagaatggat 6060
actccatcag ggataaagga ttcaattcca gcctggatca agtactcagt gcaggcccca 6120
ggagaaatac catatgatgg gatttattat gatcctcctg aagaggtctt cctttcccat 6180
atttctgtac ttacaaattt taaatatacc tttgttccta ctttaggcta gcaatttgtt 6240
tcttggaaaa taccgtttat ttcattctgg aatttcttta cacccccttt ggatccttgg 6300
aattgaattc cattctaata atagtaattt aggcaaagat caattaagct aatatggttt 6360
tatgtgaaat atatttgtat actattatta gcaatatgtg ggggatattt atgtgctaca 6420
ttttactata gaggagtgag ccgaagagcg tcttataatt tgcagagtat aaacatatta 6480
tgttgataca taaaattatt tctcatactc caccctatga atttgagata ggcttatatc 6540
attgctttgg aaagtggtaa aatgttaaat tccaagctaa atagactact ttattaagta 6600
aattccaatt cctccaaaat gaatagatct aaaagagccc cttagtgtat ttcagttagt 6660
cttatttttct gtggatctgg attttacttt tctcacttgc tactaattaa gatattttac 6720
atgcatatgc aggtaaagta tgtgttcagg catgcgcaac ctaaacgacc aaaatcattg 6780
cggatatatg aaacacatgt cggaatgagt agcccggtat gtcacctttt ccttttgatt 6840
ctttagacac tgctaggcat atgtaacacc cggccactgg ggacgtcgtc accccagggc 6900
gagtttttttt ttgactgcta gccacctgca gttttagtta attataggct tcaacgatat 6960
ttagatacta tcatcttctc acattggtct tccttatcta agcgtatgta atctaagtgt 7020
atatatttat tttcataaat gttcactttg ctactttgtt agtgaagcca gcaacattta 7080
attccaagtt attctggaat tcgtttaatg atgttgtctt gattgattga ttccaatcgg 7140
atgtttgaac ctctcaaaaa tggatgttcc aaagggagaa tttttacatt ttcttgcaaa 7200
attagtaaca attttgaact gattaatgta actttgaatg tcatttctgc tccgctatat 7260
gtttctatag ctcatagcat tatataacat ttcaaaacaa tattcatatg ttgctcatga 7320
atggtcatca atatgttata tttcttctac tttagttgat acattttgct taatatttaa 7380

aattaaattc atatttccca caattacttt gagttaaaaa aaataaacat ggccatagca 7440
cagccacagt attgttactg tcaccctaca tttgtttcta attgccatac gattttacat 7500
acatgttctt tcactgatca ttatcctttt ccggctatca ttttgttatc ttgtttgcac 7560
tagataattg cctgcattct cttgtgatta ggaaccgaag ataaacacat atgtaaactt 7620
tagggatgaa gtcctcccaa gaataaaaaa acttggatac aatgcagtgc aaataatggc 7680
aatccaagag cactcatatt atggaagctt tgggtaattt caggatccag ttttgtttgt 7740
ctttttcttc tatatccact ttatacctac tggtacctgt tgagtgtttg ggaacttcat 7800
ctttcatcgc cctcttttt agtgtcattt catgcctctt tcactgttgc ttctagtttg 7860
atagcattaa atgtacattt tggtattgtc ttcgataatc tattgtattc agatgcatgc 7920
atagtaacaa gactacacag acatatatgt aacgtcgtaa cctcaacttt atgaacatgc 7980
tctgattgga ttgatgattg catgatactg tcaatatggc attacagaaa ctcgcaagta 8040
tcagtaccaa tgcttatagt ctcctgtcat aaaaacaagt tgtttaggac aagatttagt 8100
caaatcataa aaactacaaa tatcattatc ttttaactta tttagtttga aagcatgtaa 8160
gttgacacta aaaaaagcat gtaacttata tgtgtggaat tgtattgaaa aatccttgca 8220
aatctcatgc agtcgaagga ttaagtggtc aaagatattc atcaaagacc atgcaatgtt 8280
gaatcaagta ttttgatagt aaggagtaca tgtttccaac agcatgaatt ttgatacagt 8340
atccatctgt tgagacgacg gagatgatac catccatctt aaattttagt tccgaatgat 8400
acctttttcaa ttttgacaaa caacatgaca acgtcgattt aatattttgt tacttctatt 8460
gcagatacca tgtaactaat ttttttgcgc caagtagtcg ttttggtacc ccagaagaat 8520
tgaagtcttt gattgataga gcacatgagc ttggtttgct agttctcatg gatgtggttc 8580
ataggtaatt gataaattct ttgttaataa ctcatcttcc tttaatacaa gtatcttatt 8640
gcatagttaa agctaaattt gatgatgtgt aattcatagc tccaagaaaa aggaaataca 8700
gtaagccttc atgcaaaatt gaaacattta accctcgact tactggaccc gacctgcaac 8760
taatctacaa cctctacaac atacatcaga agatgagaaa taatcgccgt aacataacaa 8820
ttaccctgac taagatatat gacaagtcaa aattttagct gtaatgttca gtgttctagg 8880
cttaagcttt aacatcagga aggaggggaa gattatccac caagtcccaa atcaagaata 8940
ttcagaaaga acctgtactg tcaaagtccc tttgatatat gaaaaaactg gacatgcgtg 9000
gggtaagaca gtcccctggc attatattaa gaagacctca cacaggtcga gaaaacatcc 9060
gaaccgtgcc acacccatac acagcggcac cgtagcccat gtgagaaacg accgcgtccg 9120
ggaccggacc ttagatatgt gctttggtat gtgatagatg aggggatttt tttaaccaca 9180
gcctgaaatt cgctcccacg gtggtagaaa cacctcggct gcgcagatta taggagtgag 9240
tatcaagttg gctggactag cgggagcgaa ggggcgccgt gctgccgctc ggctgtgaac 9300
tcaggtgagt tgcggaacac aggtgaggca gggaacaatc acttcttaga ttgatcaatg 9360
cttcccaatt gattacagac gcctgtgact ttatagtcca actctaaaca gatccttatc 9420
tcaggaactc taacagattc ttattaaaac agactcctta atctcaggat ctctaaacag 9480
actctcctta tctaaggatc ctcctaatct ttatctatac gtccctctat aatgggccga 9540
ctagagggct gtttaggcca cctactataa accataacac acggggagtc aaacccagga 9600
tctgaggagt gctactcaga ccacctaacc aactcggcta gaggcccttt cgcaagcccc 9660
tttgatatca gccacccaca tcttattttg aagtgcttga catactgttc gctgtttttt 9720
ctgctctttt tggaaccaac ttgatcaagt ttggcaccaa ttctgccatg gtcttgccat 9780
taagccagcg atcgaaccaa cattgagttc tttaaaattt agtttaacaa acacttaatt 9840
ttagtttgat tcatatttct ttcttgaaca agcaggatag ctgtgtatgt ttgtattaac 9900
agagagaaat gcaaacaacc cactttgtct tccttaggag aacaaggggg gttatggctg 9960
aggtttagag aaacgtacaa tcacctgtaa ataggataat atcaatacaa tgctacttta 10020
aactagcccc tcaacccaac ctagtcctag atcctggagt ttacaggatc cagcaaagtc 10080
ccgaagtttc agatattcta agagcacgcc atcttgcatt ttggactcaa gggaagagtt 10140
ccctcgaagg ctcacatttt tttttgtctc cataaggtcc aagcccctag aattacagcg 10200
ttgttgaaac ccttcctttg aagcatcaat ccaatccgca aagttctgct cattttgtct 10260
tggaaccaca caaccaagac caattggagg ctcacatttt ttttgtctcc ataaggtcca 10320
agctcctaga attaaagcgt tgttgaaacc cttcctttga agcatcaatc caatccgcgt 10380
cgaagctctt ctgcaagcta accgcttttc tgaagagaaa tcaactattt tctcctcgct 10440
tgtgaattcc tatagatagc agatagttct tttccctgtg ctttagttag tttccttctg 10500
gtttatgtct tgggagcgac tgctaatgct ctttgtccgt cttctttata ttaataaata 10560
tctaccacag caggggcttc tcctgctgta cattttcaaa aaaaaacata gaatgccaaa 10620
attgttaggg aaaaacacaa atagtaagaa tcactgtttt aaaggcgtcg cctaggcgtc 10680
caggcgcccc aacactccta gacgtccccg tcgcctagct aaatcagcaa gagggcagag 10740
gggaagagga aggggtggcg ccgaatcggg gccggaggcg gcgaatcggg ggcgacggcg 10800
gtgagatcag gggaggcggc gaccaaatcg tggccaacgg cagccaaacc tagggatggc 10860
aacgttttta aaacctgcgg gtagagggtt cacaaaccca cacccgcggg tctaatatta 10920

aacccgcacc agtacccatt acccgtgacg ggtataatat gatacccata ccctctaccc 10980
gcgggcatat aattatacac atatattttt atatatcatc tgtgtatgca tatatttata 11040
cacacatata aatatgcata cgcacgcacg acgcacgcac gcgggtttgc gggcacgggt 11100
acaacatttt catacccgcg agaaaaaacc cgttgggttg agaactaaac ccacacccta 11160
tggggttttt acccgcgggc acgtgggtta aatgtgtccg ttgccatccc tagccaaacc 11220
gagcgcggcg gcggcaaaat cgtgggtgac ggcggctggc aggccacatg ggctgggctc 11280
ttatgcccat acacacaaac aacacacaac agtcgaacac tcctcttctt gttaatctaa 11340
ccatagtaca atatttgttt ttattagaaa gccaataaac actttgcatt taaatagcgc 11400
ccaagatttc tacactcact tccaggtcca aaggtgagcc atggagaaag gctctatgtt 11460
gatttggaac ataaggagcc actttctcat gtttccatat atgtggatcc tgatccatgt 11520
tcatcagatt gattcatttt ctaaggacat aaagtcatag ggtttcctga ttccttttgc 11580
tacagggccc ttgtgattct ggtttctgaa tgtgtgaatt gtgaaagggc ctctagttga 11640
gttggttact caatccggtt gctgcactag caaatcagcc tactctactt tctttgaggg 11700
gtctgtcaag ttcagacctt gcagattaat ttggaggagt tgggctcctt tctgctgcaa 11760
gttctttctt tggcttgtgg ttgctgcact agcaaaccga tgttggactg cggatccgct 11820
tgccaaacga ggccttcctc acccagattc ctgctccttg tgtgactaag ccggtgagac 11880
cattcagcac attctaattg gttgtgtctt ctctagacaa atttgggtct tgactttgca 11940
atccttgcat ctggaggcca ttacgcccaa tgggacagag ttgggtttct tctcatggtg 12000
ggcgcgctct ggcaagatgg ttcccaagaa cattcgaaag gggttgaaca ctctttgtat 12060
tccgatagcc tgagaaattt ggaatttagg aacttgtgtg tgttcaaaga ctcccaacca 12120
aatgtacaac tccttctaca taggatagga agtgagggtc ttctttggtg tgctgcaggg 12180
ttgcttgagc tcctgaatag gtcgttgtcc ccagcctgct aggtgctggt ggccttggtt 12240
gtcagctttt gttgttgttt tgtgtaaaac ctagctggct agctgtttgg actagggttg 12300
ggtcattttg accttattcg ctgtcttttt ctttcgtaaa taaaatgaca cgtagctctc 12360
ctgcgtcgtt cgagaaaaaa agtactcgtc agaatagaaa attttactct ttgataaact 12420
ataactacct aatcaagaca agaatataaa attttaaata ttggaaacct atttgtcgtg 12480
gttcaaggtt ctttggcttc caagctttgg gatgagatcg attataatct gttatgtaca 12540
accttatgat tatttgaagc tctttggttt cataccttta actttgcttt tgtgttactt 12600
gcagtcatgc gtcaagtaat actctggatg ggttgaatgg ttttgatggt acagatacac 12660
attactttca cagtggtcca cgtggccatc actggatgtg ggattctcgc ctatttaact 12720
atgggaactg ggaagtacgg aacaaaaatg ctctatctcc attaatttta ttctcctatt 12780
tttctgcctg tatcgttcca acaattttat ccgtatgcag gttttaagat ttcttctctc 12840
caatgctaga tggtggctcg aggaatataa gtttgatggt ttccgttttg atggtgtgac 12900
ctccatgatg tacactcatc acggattaca agtaatttaa gctttatgcc tgttagttta 12960
tcttcacttg ctaagtctga ctggaatact ggattatgcc tgggaactag ttttgtttag 13020
tatcatattt gttatatatc attccttctt ctaatctaaa gtcatgcatt ttactttagg 13080
taacatttac ggggaacttc aatgagtatt ttggctttgc caccgatgta gatgcagtgg 13140
tttacttgat gctggtaaat gatctaattc atggacttta tcctgaggct gtaaccattg 13200
gtgaagatgt aagtgctgag tttgcttgtc atttaatatg aattctcgca tatatttgtg 13260
gaaatatttt tgtagtcgaa gttgcttttg tttatctaga caagatactc ctatttggtt 13320
atgcagaagt taatttgaat tttaatacga agtgcacact aagttactgg ttaatattgt 13380
tcttcatttc ttcaagtttc agtctatttc aactccattt ataataatgt gcttggcaag 13440
ttactgtttt aattttacat taatacacaa tacaacaaga tcatgcttta caacatgtgt 13500
gtatattaga taagtagttc atgcacatag agttgctact ttttgaagaa acatatagaa 13560
ttacttaaaa ggaactattt gaaatacatg aagaaagtta atggctgaac ctatattgta 13620
atggacagaa cactagaact ttcgggttta tactgagcct ggactataaa gaagatttct 13680
cttgaagttc aatgagttct cgacttaaat attctttcta aactactgga tggataccaa 13740
agacacaaaa aattgaaatt gtaagccact ctgctcttgt tttggcacca gatttaatag 13800
aatataaata aaaattaata atggggagaa gactcccctt ggattggctt taaagagagg 13860
agtacagaac ccttcccaac tcgtccccga accaaacttt gccaaaatcg ggatccgcct 13920
cgtcctgtcc ccagtgctcc ttggaaccaa acacatgctg agggaatata ctcctggggg 13980
atggagttat ggaccggttc caactcgtct ccgaaccaat gatttcaagt cgtccgacta 14040
atcgcgatta gtcgggctgg tcggcaatta ggacacgatt tgctgggcga ctcgactaga 14100
agacctagtc gtcctggtcg ttcgactaat cgtcgactag ggcgactagt cgttcgagtt 14160
atgtgtcctg gtcgtcccga ctagggttta gtttattggg ccttttttag cccatctaca 14220
gtctaattta gagtaccctc tcctctcctg ttctaaccta gccgccaaca gtctccacaa 14280
cctacacgtc tcctcttctc tcctcccttg gcatgccctg cagtcctgaa tcctgattgc 14340
cggcgactac accggcggct acagcaccct gcgtcccctc agggccggcc ctggacaggt 14400
gccggcggtg cggccgcacc gggcctccga aaaccagagg gcccctcccc gtgtatacgt 14460

gtatgtgtat atatatagaa aatgtttatg taaattacaa cacaaagatt gataaaaggc 14520
caaagcccat caacgtacaa tatgcatgcc tgtttaaact cgattacgta cttggcttct 14580
cgtggacgct tgagctgcaa tcctctttgc cttctcttgg aagcctccgc tcatgttgtc 14640
caggctacaa caattgttcg gcaccagcag tagcagctag cagcgacaga cgggagggat 14700
catcttgtcc gttgctctgt catctgccca aaccaggctt gccgcgcgcc gtgcctagcg 14760
gtagcgacat ggtagcgtgc cctgcctcgg taacttcgtc aagtctgcta ctctactttc 14820
gttgttcaat tttcgcctca aaccaatctt tactagttaa atgttctttt atttctgagt 14880
ataggttatt ttgtgttctg atcaccacat tccacatgca tctagtagtg aaaaaggaaa 14940
aaaacataaa gataagttca atgcatcgta gtattcatat ttttttggta tttgtgatga 15000
atggggcctt aatttttgtc tcgcaccggg ccaccaaatt ctcagggccg gccctgcgtc 15060
ccctgccccc tctcccgcag tcctgctcgc ccgctgcact gctgcagtct ccaaccgatg 15120
gtggctgcga gccagtgacg agatgagcat gtcatctgaa agtccttcct gctccccact 15180
tctcccccta aaccctaaat aaccctaaac cctaagcaac atgttacttt atttatccta 15240
gttgtttgtt ctaatttata taatgtatac aggtatattt atatatatac ctatataagt 15300
ataactacta gtctaggacg accagggacc gactaggggt cgactagtcg ccctagtcgt 15360
cgcctaatcg cgactagtcg cttggtcggt cacagagttc gactaggcga ctaggcgact 15420
tgaaatcatt gctccgaacc aaacactcag gatcaccttg tcctgtcctg gtggtccttg 15480
gaaccaaaca cacatttagg gaagtaatca tctgacctgc aagcacctgg gttcatgcag 15540
gtggatagga tctcagctag aggctctacc aaccaagtta ttgttcgatt ctcagttctc 15600
atgctccgtt gtgcaatcta atgtaaataa atgtaccacc gtgttgtacg catataccag 15660
cgtaatctcc aacaagggct gtcatgtgga atcaatagag gtctctgtca taatctgcaa 15720
cggcaacaaa ggacatacca atctttaaca cgaagagaat taactccttc ccatctgtag 15780
ttagatatta gaatattaga attgcatatt tttatgggtt cttgaaggaa gtggcaaacc 15840
tcaatattga actctaccaa cacatgtata atatgaaaaa taaagttatg gtatatttac 15900
gagtaaatat ttctatgcct tataacggga ggaaaaatat cgaattttga tattgactta 15960
ataggacatt gtcacgttat tgtcaatatt aacttatatt atggatgtca tcacaaaata 16020
actatacctt ttagttttta aagaaagggt aaaacataca tatatttgta acgtttccgt 16080
ccattttcca ttaatgtttg attgtttttt tctctcttgt tgcaacatag tgacatgttt 16140
gctagtttga caaaattagg gcaagatcat tggcttacat aaaccaatag gaaatttgaa 16200
tttatctaaa ataatttgtg agttttcttg gttattcaaa aattatataa cttgttcagg 16260
ttagtggaat gcctacattt gcccttcctg ttcacgatgg tggggtaggt tttgactatc 16320
ggatgcatat ggctgtggct gacaaatgga ttgaccttct caagtaagtg tttcatatgt 16380
atgtggacga tcattatttt gttttattgg gctgcttatt taaataatta tttttttgg 16440
cataaatgtt ataagtacat acactaatgt ttaattgcat gggccatgtg catatgtatt 16500
ttttgttgta tttgtcacca cttttgtggt ttactttaca tagtcataat gagataatgt 16560
tttatgcaca tttgcatgtg gctgcactta tgattaacga cacaaaatgt actcgagatt 16620
gatgtatatt ttcaaacttg aaactaatga acacatatga tatacattgt acactccatg 16680
ttacaaatta taagacgttt ttgcattttt agatatattt ttttactatg tatctagaca 16740
aagtgtatat ttaagtgcat tgcaaaggca atgcatctag aaaagccaaa atgcctttgc 16800
atttggaacg gaaggagtaa tattgttgca caaatcttga agttttctat gcataggata 16860
ttgcacatac cagcaatata tcagtggtgc atatatattt tgtataacta acattacacc 16920
attcactgtg ttcatatgtt ttgtgtttaa ttttctatac aaaaaatctc ttgaatctgg 16980
tactcaaatc atttgtataa gcttttgttt caaacgggca acgctctatc tgtaatatgc 17040
catctgcgta aaagacaagt atgtaatttt gtgatccttt ggttgtgtgg ttgtcggtgt 17100
aaccgggatt ctttaacctt ttatccttct tttatataat gatacacaaa ctctcctgtg 17160
cgttcgagag aaaaaatgcc ttctgcattc actttgagat atgtggtgag tttcaatttc 17220
tatttaaccg cacaggcaaa gtgatgaaac ttggaagatg ggtgatattg tgcacacact 17280
gacaaatagg aggtggttag agaagtgtgt aacttatgct gaaagtcatg atcaagcatt 17340
agtcggcgac aagactattg cgttttggtt gatggacaag gttaccctac tcattaattt 17400
tcttggtgta cttattggga catagatcat gtttcacgta ttgtttttta caatgattaa 17460
ttctatttgt ttcttcaagc tcaaggtgta ttcatggttc ccacaaaaca aatgttttat 17520
taaaagcaaa gagttagaac ttttgttagt tttctttaat ttggacttgt gtcactgttt 17580
ctctgtcatg actcatgagc attatagttg caatttcacc taagtgagtt cctgtttttg 17640
gacagttcag agtgaactac gacttattgt tttaatactt cattgagttt gtagaacaag 17700
tattggctct ttctcatcct atactttcaa aagtaggatt tggccttttt cttccactat 17760
tctttgaaaa atacatgtag aaacacggaa caataataaa tggtaacatg agaacctcac 17820
tggttctatt tatgcaggat atgtatgatt tcatggccct cgatagacct tcaactccta 17880
ccattgatcg tgggatagca ttacataaga tgattagact tatcacaatg ggtttaggag 17940
gagagggcta tcttaatttc atgggaaatg agtttggaca tcctggtgag atttaactac 18000

tttgtttcat ttaaccttcg ttgagtctta tagaacagta cctcatccaa caattatctt 18060
gcaatttatc ttttgttagt tatattagtg ttgaggactt gaggtcattt tcttcttatt 18120
attttgcaga atggatagat tttccaagag gtccgcaaag acttccaagt ggtaagttta 18180
ttccagggaa taacaacagt tatgacaaat gtcgtcgaag atttgacctg gtaaactttc 18240
tttgattgtg caaaagtcca agtttgtatt tacttttacc actgatccag tgctttaatc 18300
agcaaggtgc cattataata gttccttctt tattcatatt agcatgttcc agaagtaaaa 18360
atattactac ctttgtaaaa gttttcttta atatatgtcg ctttgtttgg tcaataattc 18420
gtcattatcg gaattgtgtt atttttacat tgtcagggtg atgcagacta tcttaggtat 18480
catggtatgc aagagtttga tcaggcaatg caacatcttg agcaaaaata tgaagtatgt 18540
tctttttta ctttttgat ttggttctgc aaggtttcca caaacatcat atttgttgtg 18600
cattctactt gtaatgtcat tttaaaaaaa atcattcctc agttttactg agcttttaag 18660
caatgaaggt ttcattatga attctttcat gttgcatcaa caactcttag gtattttcat 18720
gatcattaat agtactctgg agacagcacg ccataatggt aacgaaaaat tttctgatga 18780
aatttgctgt gtaattgcag ttcatgacat ctgatcacca gtatatttcc cggaaacatg 18840
aggaggataa ggtgattgtg ttcgaaaagg gagatttggt atttgtgttc aacttccact 18900
gcaacaacag ctattttgac taccgtattg gttgtcgaaa gcctggggtg tataaggtat 18960
gcatctatct tgcattccct atgctcaaag tgcatttctt ttcttgataa atgagttaga 19020
tatacgtact atcatgctgc aatttatcaa gtgtcattat tgatctcttt ctacggtgaa 19080
gctaggagca gctaagctgt tggtgtcagc aattcatgtt gtagttaatt taatttgctt 19140
gaaaacgtag gacgctagat ttggattttt ccaattttta ggctgcacga gcaggtaaaa 19200
ggtagcaaaa tactagggcg ccatgtttac atgtataaaa aaaacaaaac aaaaaagaac 19260
taggagttcc tgtgacggat agccgcatgc tcgttctctt ggcgtctctg atattggagc 19320
acatccgttg ttcgaaaact agccggagaa gtttctcaaa atcccactag cggaggtgct 19380
gacagcattt gctatttgta ccaggtggtc ttggactccg acgctggact atttggtgga 19440
tttagcagga tccatcacgc agccgagcac ttcaccgccg taagttttgt ggcacgtgat 19500
actgctctag gtacgcagat gtccacttgt tcctgacaga ggtgaactaa cttctgttat 19560
ggccattact tgcaggactg ttcgcatgat aataggccat attcattctc ggtttataca 19620
ccaagcagaa catgtgtcgt ctatgctcca gtggagtgat agcggggtac tcgttgctgc 19680
gcggcatgtg tggggctgtc gatgtgagga aaaaccttct tccaaaaccg gcagatgcat 19740
gcatgcatgc tacaataagg ttctgatact ttaatcgatg ctggaaagcc catgcatctc 19800
gctgcgttgt cctctctata tatttaagac cttcaaggtg tcaattaaac atagagtttt 19860
cgtttttcgc tttcctaatg cttgatggct gattgtttgc acttgtttca ttccgttggg 19920
cactgatggt cttagagtta gacaatcggc tgcagcgcat aggtttcaag ctgggggctt 19980
gcatgtccga ctacagagca gccagcaatg tggcccctgc tgcctgctct gcttactttt 20040
aaatgccacc cctcccgatt accgactcac tcagattcag acacgcaaag cactaccttt 20100
ccagtgtccc tgaagacact actcccccgt cccagcttgc cgttgcaggt atacctcggc 20160
ttgcctgcct tagatgttaa tcctacaacg atagacatgg atacggggtt ttactaatgc 20220
ttggatgcat gcattatcgt atcctcgcct cggagacatc acgcgtgcat ttggtcacac 20280
caactggtga cagggaatgt acactgagat tgctcaagag ttactcctac tcgacttgtt 20340
ggatctcacc taacgctttt caagttttta tgatacttcg ctgttggtct tggccttggg 20400
ccagtccgga gccgctctcc tgctcacatc acatgtacgt ggaatgatgt tgtctcgggt 20460
catggcatac agcttggttg gttttatttc ctccatcgat gcccaggaag cttgtgctat 20520
cattaccata agcagctaat agtatgagtt gatgagtgac aagtgacacc tgcactaaga 20580
tacatataac atcaatcaaa acaatataac aattcataca gatcaccggt gcttgcttaa 20640
aggttcaggc ttcgagtagc taggacacta gctagtactg ggctctcgat cgtttgcaaa 20700
acaatacaac agtaaaaggg agagcgagca catggaagct ggaggaataa atagatggac 20760
ggttgttaga tgcaggctac tcctactacc aatttcttta cacggcttag ggcgcaagag 20820
gcctctcctg ctctacacac tgccctagct tagctagccg tttggagggc agtgttggca 20880
gctagcagcg acgacagatg gagtgcagcg gcaaggagga gcagcagatg cagatcgtgt 20940
gcgtgcgcag cgcgtcgacg ggcggcggcg aggaggtggg ggagtgggcg gagcagtcgt 21000
cgcggtcggc gctgtcgctg ttcaaggaga aggaggagga gatcgagcgc aagaaggtgg 21060
aggtgcggga caaggtgttc tccatgctgg gccgggtgga ggaggagacc aaacgactgg 21120
ccttcatccg ccaggagctg gagctcatgg ccgaccccac gcgccgggag gtcgacgcca 21180
tccgcaagcg catcgacaag gtcaacaggc agctcaagcc gctcggcaag acgtgcctca 21240
aaaaggtacg tcactacaag acactggaac acaacacgag tgtttctcca ttcgggtttt 21300
tcattcgacc gaatggctat agaacgaagt caagcaggta gcaacacgta ctgcttgctg 21360
atcgtgtgtt tgtctccgaa catgtgcgtg tgcaggagaa ggagtacaag atgtgcctcg 21420
aggcctacaa cgagaagaac aacgagaaag ccacgctggt caataggctg ttggaggtca 21480
gttcaattct tttgcaatcc gcgcgcgcta tagctagctt ttgttttgag aaccgtctcg 21540

gtaggcatat tttgttctgc tgcggaaaat aaaaaggatt aaggctgttt tagaaccaat 21600
aataattcct gtgaaattca tataccaatc cgatggatgg aattggttgg gccatttcca 21660
tattattact gtactactag gcccaactga cttggtcctc tatttttttcc ctaataagaa 21720
cgtgtctgtt gacacatgtt tcactcagag gttggtcatc taaaacaata aaataataac 21780
gtgttctttc tggcagctcg taagcgagag cgagaggctg cggatgaaga agctggagga 21840
gctcaacaag acgatagaat ccctctacta gatctgatcc ttgggtcttg ggcttcaaag 21900
aagcaaaatt ccgttgctct ttctaccgcc gaagccgagt acattgctgc aggacattgt 21960
tgcgcgcaac tactttggat gaggcaaacc cttagggact acggttacaa attaaccaaa 22020
gttcctcttc tatgtgataa tgagagtgca atcaaaatgg cggataatcc cgtcgaacat 22080
agccgcacta aacacatagc cattcggtat cacttttaa gggatcacca acaaaaggga 22140
gatatcgaga tttcatacat taatactaaa gatcaattag ccgatatctt taccaagcct 22200
cttgatgaac aaacttttaa caaacttagg catgagctaa atattcttga ttctaggaac 22260
ttcttttgtt gatttgcaca catagctcat ttatatactt ttgatcatgt ctctttcata 22320
tatgctatga ctaatgtgtt ttcaagtgaa tttcaaacca agtcataggt gtattgaaag 22380
ggaattggag tcttcggcga agacaaaggt tccactccgt aactcatcct tcaccatcat 22440
ccgagcaatc tccatctttg gggagaaaag gactccatct ttggtataat cttcactcat 22500
ttatttatga ccaaaggggg agagagtaat tcaagggctc taatgactcc ttcttggcga 22560
ttcatgccaa agggggagga agtattagcc caaagcaaaa ggaccgcacc accacctaat 22620
tttaaaagtt ttttttcaat tggtatgatt ctcaattgat aaatttcaaa ttggtatctt 22680
attgtgttca aaaggggaa aaaagtagtt tttcaaaatt gatatgacaa caccctcttg 22740
aacactaaga ggagaatttc atctagggga gctttgttta gtcaaaggaa aagcatttga 22800
aacaggggga gaaaatttca aatcttgaaa atgcttcgca aaatcttatt catttacctt 22860
tgactatttt gcaaaagaac tttgaaaagg atttacaaaa agaatttgca aaaacaaaac 22920
atgtggtgca agcgtggtcc aaaatgttaa aaatgaagaa acaatccatg catatcttat 22980
aagtatttat attggctcaa ttccaagcaa cctttgcact tacattgtga aaactagttc 23040
aattatgcac ttccatattt gctttggttt gtgttggcat caatcaccaa aaaggggaag 23100
attgaaaggg aattaggctt acacctagtt cctaattgat tttggtggtt gaattgtcca 23160
acacaaataa ttggactaac tagtttgctc tagtgtataa gttatacagg tgccaaaggt 23220
tcacacttag ccaataaaaa gaccaagaac taggttcaac aaaaagagca aagggataac 23280
cgaagtgtgc ccagtgttcg gtgcaccagg ggacttcacg ctgaactcct caccttcggg 23340
aaaatccaga ggcgcttcgc tataattcac cggactgtcc ggtgtacacc ggacagtgtc 23400
cggtgctcca aggaagggcg cctcaggaac tcgccagctt cgggaattc 23449

<210> 3
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> /note = "synthetic construct"

<400> 3
gactgaattc ctgcgcagga ggcagagctt 30

<210> 4
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> /note = "synthetic construct"

<400> 4
gatcgaattc catagatacg tggagcagca 30

<210> 5
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> /note = "synthetic construct"

<400> 5
ggcgacacga ggcacagcat 20

<210> 6
<211> 39
<212> DNA
<213> Artificial Sequence

<220>
<223> /note = "synthetic construct"

<400> 6
ggtcgactcg agtcgacatc gatttttttt ttttttttt 39

<210> 7
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> /note = "synthetic construct"

<400> 7
gactgagctc ataccaaatg aagccaggag 30

<210> 8
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> /note = "synthetic construct"

<400> 8
ggtcgactcg agtcgacatc ga 22

<210> 9
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> /note = "synthetic construct"

<400> 9
ccagctccac ggttgttcgt gt 22

<210> 10
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> /note = "synthetic construct"

<400> 10

cgatggatcc tgtgacggcg tgtgagtccc 30

<210> 11
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> /note = "synthetic construct"

<400> 11
gatcggatcg aactgatcag 20

<210> 12
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> /note = "synthetic construct"

<400> 12
cctaattgta gccctgcagt ca 22

<210> 13
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> /note = "synthetic construct"

<400> 13
gactggatcc tcgccttcgc agccggatcg 30

<210> 14
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> /note = "synthetic construct"

<400> 14
gaaagtcgac gaagagagaa tgaaagcgaa 30

<210> 15
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> /note = "synthetic construct"

<400> 15
gcgcgggtcc gtcgtgcctt tt 22

<210> 16
<211> 30

<212> DNA
<213> Artificial Sequence

<220>
<223> /note = "synthetic construct"

<400> 16
gatagtcgac cgacgcgcaa cggcctgcct 30

<210> 17
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> /note = "synthetic construct"

<400> 17
gactggatcc tcgccttcgc agccggatcg 30

<210> 18
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> /note = "synthetic construct"

<400> 18
gatcgtcgac cgctcgtctc cgtcctatat 30